THE CONTROL OF MANUAL ENTRY ACCURACY IN MANAGEMENT/ENGINEERING INFORMATION SYSTEMS

Final Technical Report (Phase I) Prepared For

The National Aeronautics and Space Administration
George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812
Under Contract No. NAS8-37311

Authors

Daniel Hays
Henry Nocke
Harold Wilson
John Woo, Jr.
June Woo

(NASA-CR-179171) THE CONTROL OF MANUAL ENTRY ACCURACY IN MANAGEMENT/ENGINEERING INFORMATION SYSTEMS, PHASE 1 Final Technical Report (Gamma Research) 68 p Avail: NTIS HC A04/MF A01 CSCL 05B G3/82

N87-29370

Unclas
0097807

REVIEWED AND APPROVED BY

John Woo, Jr.

John Woo, Jr.
Principal Investigator
and President

GRI-TR-87-11
August 28,1987
GAMMA RESEARCH, INC.
904 Bob Wallace Ave.
Suite 124
Huntsville, AL 35801
(205) 533-7103

TABLE OF CONTENTS

ATTACHMENT B
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

PHASE I SUMMARY

TO BE COMPLETED BY OFFEROR

SBIR 86-1

PROPOSAL NUMBER
07 06 7103

NAME AND ADDRESS OF OFFEROR

GAMMA Research, Inc.
904 Bob Wallace Avenue, Suite 124
Huntsville, AL 35801
(205) 533-7103

PRINCIPAL INVESTIGATOR

John Woo, Jr.

TITLE OF PROJECT

Control of Manual Entry Accuracy in Management/ Engineering Systems

TECHNICAL ABSTRACT (LIMIT 200 WORDS; INCL. STATEMENT ON PHASE I RESULTS)

A growing national need exists for the development of a computer work station technology which is interactive and assists the user in controlling the accuracy of manual data entry tasks during data base preparation and during communication over local area networks. Such a technology would affect increasing numbers of NASA employees at all levels as computer terminals become widely used in management, engineering, and accounting for decision-making, project and program control, and intra-agency communications. Data flow in networks produced by manual data entry at controlled levels of accuracy would impact the management and technical performance of projects and programs and would have a favorable impact on productivity.

The Phase I research effort successfully demonstrated that clerical personnel can be tested for proofreading performance under simulated industrial conditions. A statistical study showed that errors in proofreading follow an extreme value probability theory. The study showed that innovative man/machine interfaces can be developed to improve and control accuracy during data entry.

POTENTIAL COMMERCIAL APPLICATION OF INNOVATION AND THIS PROJECT

This research is directed at improving the methods and equipment used in manual data entry into computer data bases so that accuracy and speed are increased. Such methods and equipment would have universal application commercially with the increasing availability of computers and the trend toward development of distributed networks where data entry is performed by end users who may not be highly skilled in keyboard operations.

FOR NASA USE ONLY

AWARD NUMBER

Section A

INTRODUCTION

This Final Report presents Gamma Research's description of the Phase I research effort to meet the specific technical objectives of this contract.

The Phase I Technical Objectives have been completely fulfilled by Gamma Research. All line items proposed in the Phase I Work Plan were successfully completed. The research has shown that the VODEM experiments with clerical personnel can provide a significant improvement in speed and accuracy over conventional proofreading in industry and commerce. The foundation has been laid for the Phase II development of the hardware and software means (VODEM Technology) to control the accuracy of manual data entry tasks.

A.1 Benefits to NASA

Manual data entry is a necessary part of any large enterprise engaged in business and technical analyses. The benefits to NASA would be increased accuracy and control of data over the whole spectrum of operations, including financial, payroll, contracts, engineering analysis, and the extensive documentation required in many of these areas. The development of such tools would be timely, considering the trends away from a centralized data processing facility and the growth of end user terminals.

A.1.1 Application to the Space Station

The development of the VODEM Technology would add another dimension to NASA's interest in controlling quality in all aspects of operations. Gamma Research, Inc. (GRI) has identified important applications of the VODEM Technology to NASA's mission of development of the Space Station:

1). For control of data entry accuracy among ground controllers at consoles while they are going through several cycles of proofreading,

2). For evaluation of astronauts for data entry accuracy while in orbit.

SECTION B

Experimental Plan

The design for data gathering attempted to provide some of the realism of a work setting while maintaining control of experimental materials and conditions. A problem with many psychological studies of text correction has been the short period of time - often around half an hour - that the participants corrected textual errors. Field studies of text have typically, by contrast, involved real situations and long periods of time, but have had little control over conditions or even kinds of text presented in some instances.

Having subjects work for five half-day sessions seemed to balance realism against fatigue. It was also a practical length of time in most cases to arrange for 'borrowed' personnel who were used in Study A along with paid participants.

A minimal set of experimental conditions for Phase I investigation had to include the following:

- the Vodem versus side-to-side (STS) presentation,
- words (high meaning units) versus serial numbers (low-association material),
- 3 to 5 levels of density of embedded textual errors,

all suitably controlled for distribution of errors, error type, external conditions, and so on.

This 2 × 2 × 3 or 2 × 2 × 5 design includes 12 or 20 experimental condition cells, respectively. One participant working five half-days can fill only a small percentage of these. Taking another design approach, that of replicating the matrix using many subjects, would have been prohibitive. The solution taken was to have basically a "subject-as-own-control" design, but to assign the limited number of participants over treatments so as to balance sequence effects, and to have enough within-subject comparisons for meaningful results.

It is important to note that this design strategy is biased toward comparisons of a given person's performance under one condition of testing with his or her own performance under other conditions. We do present grouped data from Study A for varying sets of subjects assigned to given conditions, but doing so is based on indications of the comparability of these participants, and should be interpreted carefully.

Study B, a leaner and more balanced design, was constructed to make essential comparisons on either a group or within-individual basis. It is described below.

B.1 Design of the Two Studies

Study A was the more comprehensive of the two studies in terms of number of values of variables included. Its design is outlined in Figure B.1. One module of three participants allows comparisons for two or three values of error density for STS and Vodem presentation, for Random Words. A parallel module of three

Participant/Condition

Random Words

Day	A	B	I
1	MID - STS	MID - VOD	HI - STS
2	LO - STS	HI - STS	MID - VOD
3	MID - VOD	HI - VOD	MID - STS
4	HI - VOD	MID - STS	LO - STS
5	HI - STS	LO - VOD	HI - STS

Serial Numbers

Day	C	D	J
1	HI - VOD	MID - VOD	HI - VOD
2	HI - STS	LO - STS	MID - STS
3	MID - STS	LO - VOD	HI - STS
4	MID - VOD	HI - VOD	LO - STS
5	LO - STS	HI - STS	MID - VOD

Words and Numbers

Day	E	F
1	Words - MID - STS	Words - MID - VOD
2	Serial - MID - STS	Words - HI - STS
3	Serial - MID - VOD	Words - HI - VOD
4	Words - MID - VOD	Words - LO - VOD
5	Misc. Conditions	Misc. Conditions

All Five Error Densities - Morning and afternoon sessions for five days, with embedded error densities L1 (LO), L2, L3 (MID), L4, and L5 (HI). Subject **G** Random Words. Subject **H** Serial Numbers.

Figure B.1 - Design for Study A

participants allows such comparisons for Serial Numbers. A two participant module allows comparisons between Random Words and Serial Numbers, with an error density comparison for one or the other type of stimulus. A last module, for temporary hires working for ten sessions rather than the usual five, allows STS and Random Words to be compared at five levels of embedded error density, the HI, MID, and LO of the rest of the design, and two intermediate levels, L2 and L4.

Participants for this study were primarily recruited from cooperating agencies and corporations, with two from one agency, two each from two private sector organizations that had relationships with the contracting agency, and another from another such company. In addition, two persons were hired for 10 sessions each from a part-time help firm (one dropped out immediately). Finally, three college students were hired to augment the sample.

Of the eleven participants in Study A, 8 were female and 3 male.

Level of experience of the agency and most corporate subjects was fairly high in typing, data entry, and associated text-processing areas, ranging from several months to 11 years. Average experience with office work was 6.25 years. (Each participant filled out a questionnaire listing experience in several text-related areas, schooling and job training, visual acuity, and other factors, to provide this information.) Proofreading experience as such was fairly well represented in the corporate and agency participants, with seven having experience from 1 to 10 years in at least occasional proof-reading, and three indicated "daily or almost daily" proofreading experience for periods of one to three years. One work organization participant listed no proofreading experience. The college students and temporary hire showed little formal experience in data entry, statistical typing, or proofreading, though all were familiar with the use of personal computers for word processing. Background experience is not presented in detail here for individual subjects. The main point for present purposes is that although levels of experience were fairly varied, this group was familiar with general text processing, though of course not with the specific presentation techniques employed in this investigation.

Implementation of Study A. Because of various practical contingencies, the Study A design was implemented unevenly. One temporary hire dropped out within the first hour, and the decision was made to have the other correct Random Words. Two participants, borrowed from a cooperating organization, were called away after completing only three of the five sessions. One participant developed a neck problem during a side to side condition, and was switched to Vodem presentation and let out early. One session for one participant ended in machine failure, and there were a few problems in the availability of material on functioning fixed disk drives of the PC's.

The net effect was more data on Random Words than on Serial Numbers. In a sense, the Serial Numbers provide more of a "pure" case of information processing, because of their lower association level. A person can compare a misspelled word against an internal, memory representation more easily than he or she can compare a Serial Number, which is quite arbitrary.

Study B. Though Study A produced a fairly rich body of individual data, explored in some detail in Section D below, it exhibited a fair amount of variability. Because of its logic of design, and the practicalities of implementation, interpretation of grouped data statistics has to be done with care. For these reasons, a much simpler study, focusing on Serial Numbers, and having only the middle and low levels of embedded errors, was run during the last month of the project.

In this study, four persons from a temporary help firm proofread Serial Number material at two levels of error density. Distinctive features of this study, which profited from the experience of Study A, included the following:

- a more extensive training session the first day,

- elaborate feedback, with examples, and comparative statistics, to each participant after the first day's training,

• less elaborate feedback at the end of each half-day's session on how the person had done that day,

• a more homogeneous subject group relative to experience level,

• comparable control of motivation factors stemming from the conditions of payment (rather than being 'on loan', etc.),

• control of certain factors of stimulus detail that had turned out to result in somewhat misleading error-correction figures in Study A.

The design of Study B is shown in Figure B.2.

Session	All Subjects
0	Training
1	Serial - MID - VOD
2	Serial - LO - VOD
3	Serial - LO - STS
4	Serial - MID - STS
5	Serial - MID - VOD repeat

Figure B.2 - Design for Study B

Participants in Study B had less job experience related to text processing or office work than did the corporate/agency participants from Study A, but more than the college students in that Study. Average length of office experience for Study B was 16 months. Experience in data entry, statistical typing, and working with personal computers was also sparse. However, 3 of the 4 indicated "occasional proofreading" for periods ranging from 3 to 15 years. Three were female; one, male.

B.2 Conditions of Testing

All sessions were held in the MSFC Micro-computer Training Facility. IBM PC-XT and IBM PC-AT computers were equipped with Super Screen™ anti-glare screens,. For STS conditions, Rubbermaid Copy Holder™ copy stands with moveable line indicators were used. Because of availability, only XT's were used in Study B.

Response times of the machines with the programs used were the same. Screens were identical. Keyboards did differ. However, participants had indicated preference for keyboards on their initial questionaire. and the only two persons who indicated a preference for a standard AT keyboard were assigned to those machines.

Uniform tables and chairs were used. (The chairs had arms, which many participants complained of as annoying. Perhaps at a future time this facility might be equipped with chairs more suitable for computer use.) Ambient noise level in the facility was high enough to mask incidental noises. Overhead lighting was reduced during the actual proofreading sessions to diminish remaining glare.

Section C

COMPUTER PROGRAM DEVELOPMENT

C.1 Computer Language Requirements

For the programming demands of the Phase I contract, a decision was made to develop and convert all computer programs from IBM BASIC to the new Microsoft QuickBASIC. IBM BASIC consisted of an interpreter that must convert each line of computer coding to machine language before that line can be executed. This resulted in slow performance in multi-line complex programs. In work leading up to this contract, the need for improved performance was satisfied with the use of the IBM BASIC compiler. The compiler converted all lines of coding to machine language to form a new stand-alone executable file. A definite increase in execution speed was realized, but all other limitations still existed. The compiling process lasted up to forty minutes, involving ASCII conversion, compiling, and linking. The compiler was internally limited to a relatively small amount of working memory, causing some problems with the RAM-intensive proofreading routines. There were many primitive features in the IBM BASIC when compared to other available user-friendly languages and utilities where many have word processing-like editors, mouse supported pop-up menus, large working memory, and advanced error debugging. These limitations bought support for the purchase of a new high-level language to increase productivity.

Microsoft QuickBASIC version 2.0 met almost all of the requirements that were set forth. It included advanced editing, compiling, debugging, and mouse support. Microsoft patterned this new language to rival other high-level languages like PASCAL and C, while still providing the programming ease of BASIC. Features like 100% compatibility with IBM BASIC, a ten times performance increase over IBM BASIC, and automatic in-memory compilation were main contributing factors in the final decision to make Microsoft QuickBASIC the language for the contract. A Mouse Systems PC Mouse was also bought to take advantage of the mouse support that greatly extends cursor control and the ease of operation during interaction with the program. This eliminates complex multiple key stroke commands, replacing them by hand-moving a small box and pressing a button to choose any command.

The first major project concerned grasping many new features in the QuickBASIC editor and understanding the operation of the automatic compiler and of the new programming elements. The most obvious attraction was time-saving mouse support that was most important during the early half of the contract when old programs were modified to QuickBASIC, where convenient interaction with QuickBASIC proved to be of great importance. The ease of operation that the mouse provided to QuickBASIC was transplanted to a few programs in the developing stage when the mouse was seen as a new tool for the proofreading programs. QuickBASIC also intrigued us and allowed us to create other performance raising routines that were practical and extremely proficient. Many programs were substantially reworked. Most fit in the category of file generation because of the considerable amount of calculations which must be continually cycled to create a file of considerable length. All of the programs benefited from QuickBASIC's easy error correction and modification abilities.

C.2 Computer Programming

In developing the computer programs, the general interest was to provide a streamlined approach to gathering the proofreading data on each individual that was tested. Computer random generation techniques were employed to supply the most efficient means for creating the proofreading data files with embedded errors. This allowed unbiased, random data to be produced at a high speed. The PRF proofreading program tested all the individuals on the proofreading data files and recorded the results. Some programs provided a partial analysis of the results by printing them out with respect to the file that was proofread, summarizing results, comparing the different aspects of the results, and plotting the results against each other. The computer played a major role toward the processing of data to give full view of the proofreading results for human analysis. The following paragraphs will describe the important programs that were used during the contract.

C.2.1 Proofreading Data File Preparation

Four steps are needed to prepare a data file for proofreading. In the first step, continuous lines of random text are generated. There are two programs designed for this task. For word generation, the program RANDWORD randomly reads in words from a 50,000+ word dictionary to compose lines about 75 characters in length. The only user selectable parameter is the number of lines that will be created. Next, the program SERIALQB formulates text files filled with serial combinations like license plate numbers, telephone numbers, and ID codes. The user selects the number of fields (serial 'words') per line and the number of characters per field. Each character in the field is assigned a set from which that character is randomly selected. A set can consist of only letters, only numbers, letters and numbers, or a special character like a dash which is used in telephone numbers. For example, a license plate number is chosen to have seven characters. The first three characters are assigned to set A, a set of only letters. The last four characters are assigned to set B, a set of only numbers. This brings about three random letters and then four random numbers in this character field. In addition, different types of serial combinations may be alterated every 25 lines (or any other number of lines to produce a variety of data for testing). The purpose of both data generation programs is to quickly create multiple lines of text for processing in the next step.

The second step adds embedded errors to lines of text in a random but controlled procedure. The most important feature is the user selectable error densities. There are up to seven different types of errors that can be implanted in the text. Each error is given a value for its frequency of occurrence. This method allows high density errors, low density errors, favoring of certain errors over other errors, and even the elimination of a type of error. RED is the basic program that is used in the second step, but there are ten options for RED. The options have different error densities. Five of the options are for word files, and the other five options are for serial files since the serial file options must eliminate the upper-lower case errors. The five files of the two sets have five different error densities that range from extra low error densities to extra high error densities. In the construction of each of the files from RED, three lines are outputted for every one line inputted. Output line one (PDP(1)) is identical to the input line. Output line two (PDP(2)) is the input line with any embedded error added. Output line three is the control line which contains the information about any errors embedded in output line two. The information is instrumental in all future use of the output files from RED. Step two concludes most of the construction of data that is to be used in proofreading.

For the third step, a special version of the PRF proofreading program was created, that was called SCAN. This program processes the proofreading data files as if an individual was to constantly go to a new line without any correcting. An output file is made comprising of all the error information since no errors were corrected. The data in the output file gives an actual accounting of the error distribution so the file can be determined to be acceptable in its distribution of errors or to be rejected for undesirable error density, like a too low error density percentage in the RED program during error embedding.

In step four, the data that was outputted in step three must be translated to an easily understood form. The MAYOUT program had already been used in previous projects to examine data that was outputted by the PRF proofreading program. MAYOUT was modified to produce a new version called MAYSCAN that discarded many user options, replacing them with a default since all data run through MAYSCAN would be of a similar nature. MAYSCAN automatically interpretted the output data from SCAN into informaion summaries every 100 lines. A data table for each 100 lines with multiple summary tables was printed out on paper to provide easily accessible, permanent records on every error in the proofreading data files. The data files could now be deemed acceptable with the information included in the data tables. The only data file that was rejected resulted from a low miscalculation of the missing line error density. Steps three and four insured error densities and error distributions that were consistent with the goals of the proofreading testing.

Step five applies only to the proofreading data files which are to be used in side-to-side proofreading. Because only one line of the data (PDP(2)) is put on the screen for side-to-side proofreading, there was a need to print out the other line (PDP(1)) for comparison against the line on the screen. A program had been developed that read proofreading data files and only outputted to the printer the first line (PDP(1)) of the two line pair of lines to be proofread against each other. Each page in the printouts contained 25 lines of double spaced text. After the conclusion of step five, the proofreading data files can now be used in the PRF proofreading program.

To satisfy the requirement for proofreading data, ten files were created as word files, and ten files were created as serial files. In the ten files, there were two sets of five files. One set would be for VODEM testing. The other set would be for side-to-side testing. The files in each set differed from low error density to high error density. Each file contained 1800 lines or about 300,000 bytes of data. These twenty files and two extra middle error density files constituted all of the data that was used for proofreading testing in the duration of the contract.

C.2.2 PRF Proofreading Program

The PRF proofreading program tested individuals under controlled conditions for accuracy in error correction. The PRF7 computer program was modified from our standard proofreading program, PRF33. PRF7 was used as the proofreading testing program throughout the contract. In PRF7, if the final line (PDP(3)) that is corrected by the user through data entry keystrokes contains any uncorrected errors, the final line is saved as a new data record in the user's output file. This was the most significant change in the basis of the program. Other modifications derived from the conversion of PRF7 to a new language, Microsoft QuickBASIC.

An additional version of the PRF proofreading program, PRFMOUSE, was created to take advantage of the ease of operation that was offered by the PC Mouse. The task of developing a user-friendly environment with easy to comprehend menus and continuous user control was a time consuming process. PRFMOUSE was not completed until the latter stage of the contract. PRFMOUSE only used the keyboard for data entry corrections. All options and commands were displayed on the screen and selected through the movements of the mouse and the pressing of the mouse's buttons. Even the selecting of the position of errors was controlled by the mouse. This made it natural for the eyes to continue looking at the screen even while commands are being selected because in order to view the actions of the mouse, a person has to look at the screen. This eliminates multiple head movements between the screen and the keyboard to select and verify commands that occur when using the keyboard. Still, a person must go through the keyboard process when typing in data entry corrections. By allowing the eyes to stay affixed to the screen, the speed and accuracy of the selection and verification of commands are increased. Overall, time is saved in a variety of areas, and the ease of operation is improved.

CARDA is a supplemental program to the PRF proofreading programs. It performs the task of providing the users of PRF with the information to run the proofreading program. The information includes the user's code name, the date, the computer number, the disk number, and the commands to start the proofreading program. The program is also designed to create the statistical files for the storing the user's proofreading results. Below is a typical example of the information that is provided to the user in order to run the proofreading program:

```
INCH     MONDAY 29    COMPUTER NO. 10    DISK NO. 98

A>
A>C:PRF7

ENTER YOUR PROOFREADING TEXT FILE ? C:FLOPPY10.W05
ENTER YOUR OUTPUT STATISTIC  FILE ? A:INCH-F10.W05
ENTER YOUR HEADER RECORD ? KAREN Baker
PRESS THE [RETURN] KEY TO START THE PROOFREADING
```

C.2.3 Computer Analysis Programs

The MAYOUT3 computer program analyzes the proofreading results from PRF7 into a line-by-line format of error and proofreading information with comprehensive summaries. MAYOUT3 was modified to accommodate the new version of the PRF7 proofreading program. The format for the input data file of the MAYOUT3 program was changed such that an additional data record of the proofreading line (PDP(3)) could be read from the PRF output file.

The SAMP9 computer program is designed to analyze the results of the proofreading tests up to the specified number of the embedded errors per sample. Each sample contains 25, 50, 100, or more embedded errors from the original proofreading file. SAMP9 reads this proofreading file and the line-by-line results from the PRF proofreading tests. It then outputs various statistical results for each sample. These statistical results per sample include the following:

1. residual error
2. number of characters and keystrokes
3. time to complete the sample
4. correction time
5. % ratio of corrected to original errors
6. ratio of the time to complete the sample to the number of characters in the sample
7. % ratio of residual error to sample size
8. ratio of sample time to average sample time
9. error density (errors / # of characters)
10. user processing time per character

SAMP9 also saves the statistical results by creating two random access files. One is for embedded errors, and the other is for the corrected errors. These results include the number of characters, length of time, number of keystrokes, and correction time between two adjacent errors. In the subsequent run, these two random access files can be accessed again without regenerating these statistical results to save time. Then, SAMP9 can immediately compute the final sample results for various sample sizes. Also, a third random access file is created to save the final results for each sample. SAMP9 can use this file to plot X and Y points on the printer.

The SPLOT2 computer program is a plotting program to obtain a scatter plot on the printer for any two functions of the sample data being used as X and Y points. This program also can obtain the histogram plot for a specified function. SPLOT2 reads the random access file that is generated by the SAMP9 program, and it outputs the various sample function plots. For each plot, SPLOT2 allows the user to select a plot type, either a histogram or scatter plot, and also a printing mode, either on the screen or on the printer. All of the sample data functions are listed in the menus, and the user can select any one or two functions for a specified plot. The compressed print mode is used to allow 136 columns of printing on the printer. For the printer output, 130 columns are used to plot the X points, and 87 rows are used to plot the Y points. The first and last three columns and rows are used for the two X and Y axes. These axes are made with plus signs and are labeled with the number of columns or rows. For the screen display, 74 columns are used for plotting X points. The data points can be printed as the numbers from 1 to 9 and the letters for A to Z to indicate a difference between the data points.

The POINTPRT computer graphics program is a general purpose bitmap display program. This program also has an option to print the bitmap on the printer in the graphics mode. The input of this program is the X and Y points of any function. This program computes X and Y scale factors and the interval size between tick marks according to the input data points. The origin of the X and Y coordinates is set as (Ø,Ø) and is located in the left corner of the screen. The X axis is drawn on the right side of the origin, and the Y axis is drawn on the top side of the origin. Each axis has 1Ø tick marks, and these tick marks are labeled as increasing positive numbers for the X and Y values. The title of the bitmap and the X and Y axis are also displayed on the screen.

C.3 Example of the PRF Display Screens

When a user begins to proofread a data line pair, the PRF display screen appears as shown in Figure C-1(a) Proofreading Computer Screen. The top two lines indicate character positions, numbered from Ø1 to 8Ø. The third line is PDP(1) which the user assumes to be correct. The fourth line is PDP(2) which the user is to proofread for errors by comparing it with PDP(1). The next two lines instruct the user to use the function key [F9] for a new PDP data line pair or to use the function key [F1Ø] if a correction needs to be made. Figure C-1(b) shows the screen after corrections have been completed. The instructions appear below the PDP data line pair. A correction mode is chosen between replacement, insertion, and deletion. Replacement is the automatic default, and insertion and deletion are selected by the function keys [F7] and [F8], respectively. A two-digit character position is entered by the user as 'Ø1' which appears on the 11th line. Accordingly, an asterisk appears at character position 'Ø1' on line 5. During correction by the user, the characters being entered appear on line 6 below the asterisk. After the return key is pressed, the characters that were entered are copied to line 4 (PDP(2)) at position 'Ø1' in a highlighted mode. Now, the lines PDP(1) and PDP(2) are considered to be identical so the function key [F9] is depressed to allow the next PDP data line pair to be displayed.

C.4 Notation

C.4.1 Abbreviations and Definitions

STS	–	Side-to-side proofreading
V , VOD	–	Vodem , head-on computer aided mode

C.4.2 Error Density Abbreviations

L1	–	8Ø errors per 5Ø,ØØØ characters
L2	–	125 errors per 5Ø,ØØØ characters
L3	–	1ØØØ errors per 5Ø,ØØØ characters
L4	–	18ØØ errors per 5Ø,ØØØ characters
L5	–	26ØØ errors per 5Ø,ØØØ characters

```
00000000011111111112222222222333333333344444444445555555555666666666677777777778
12345678901234567890123456789012345678901234567890123456789012345678901234567890
HENS NOMINALLY POLYNESIAN OK INNOVATE FELT BUNGLER PICKLES SMUDGING UNWISELY
ENS  NOMINALLY POLYNESIAN OK INNOVATE FELT BUNGLER PICKLES SMUDGING UNWISELY


         <F9 >   NEW LINE
         <F10>  CORRECTION







         ****PROOFREADING  MODE****
```

Figure C-1(a). Proofreading Computer Screen.

```
00000000011111111112222222222333333333344444444445555555555666666666677777777778
12345678901234567890123456789012345678901234567890123456789012345678901234567890
HENS NOMINALLY POLYNESIAN OK INNOVATE FELT BUNGLER PICKLES SMUDGING UNWISELY
HENS NOMINALLY POLYNESIAN OK INNOVATE FELT BUNGLER PICKLES SMUDGING UNWISELY
●
H
<F9> NEW LINE    *** ENTER 99 FOR TERMINATION , ENTER 88 FOR BREAK ***
<  > REPLACE ONE OR MORE CHARACTERS
<F7> INSERT A CHARACTER
<F8> DELETE A CHARACTER

         ENTER 2 DIGIT CHARACTER POSITION FOR CORRECTION.   01



         -----CORRECTION  MODE-----
```

Figure C-1(b). Completed Proofeading Screen.

C.4.3 The Eight Error Types

I. Substitution of one character for another
II. Transposition of two adjacent characters
III. Missing character
IV. Added character
V. Upper/lower case error
VI. Missing word
VII. Missing line

Section D

EXPERIMENTAL RESULTS

Gamma Research, Inc. completed the Phase I experimental program to evaluate clerical personnel for proofreading ability under simulated VODEM and industrial conditions with several error densities. Table D-1 summarizes the findings for the 15 subjects. The group for Study A is noted as Subject(1) - Subject(9) plus Subject(A) - Subject(B), for a total of 11 subjects. Study B included Subject(C) - Subject(F), for a total of 4 subjects.

The individual average total time per line, T_L, for all subjects in Study A was plotted as a function of error density, $\bar{N}_o$. Then the points were fitted with a linear regression line. The results are shown in Figure D-1 for the four combinations of Words-Serial and STS-VODEM. Also shown is the coefficient of correlation for each fit. Figure D-2 shows the Performance Charts for Words at the STS and VODEM imaging modes, averaged over Study A. Figure D-3 shows the average performance of Study A for Serial. Due to the various strategies employed by the subjects for proofreading Words, Figure D-2 does not show an improvement for the VODEM mode at all error densities tested. However, Figure D-3 shows clearly the performance improvement of the VODEM at three error densities: L1, L3, and L5. Both figures also show the gains in performance with increasing error density for both imaging modes. Figure D-1 shows the gain in speed of total time per line for both imaging modes.

Figure D-4 showns the Performance Charts for the subjects in Study B for Serial Numbers, averaged over the four subjects. A clear advantage is seen for using the VODEM imaging mode. Also shown are the gains in performance with increasing error density for each mode. Figure D-5 shows an individual from Study B with the greatest gain in accuracy using the VODEM and increasing error density over the STS mode. Using the symbol chart from Table D-1, Subject(F) demonstrates the high accuracy that can be obtained. Also shown is the performance of Subject(C) who had the highest error level. In this case, the VODEM interface was able to increase the accuracy level over STS by a large factor (5.7).

Using the Error Recovery Method that was discussed in detail in the Phase I proposal, it is possible to calculate the job time function for individuals and group averages. Using Figures D-1, D-2, and D-3 the error level for the same amount of proofreading time for 25,000 original errors in five million characters was calculated for the average of Study A. It was found that STS without embedded errors (the industrial method) was less accurate than VODEM with embedded errors. The gain in accuracy for VODEM was 7.7 using serial numbers. Only two individuals could be self-compared because of the lack of data. Subject(4) had a gain in accuracy of 19.7 using serial numbers. Subject(6) had a gain of 10.2 for serial numbers.

For Words, the group average performance for Study A gave a gain in overall performance of 15.6 when the VODEM is used over STS. Only one person had enough data to calculate an individual gain. Subject(3) had a gain of 3.5 in accuracy.

Table D-1a. Data from the Proofreading Experiments

ject	Plotting Symbol	Number of Lines Proofread	Average time (sec) for Zero Error Lines, T_z	Average Time (sec) Over All Lines, T_L	Percent Residual Error C	File Name	Imaging Mode		Error Density				
							STS	VODEM	L1	L2	L3	L4	L5
		220	20.6	30.7	8.4	1-F08a	X				X		
		251	24.4	42.4	4.4	1-F10	X						X
		417	12.6	26.5	5.4	1-F05		X					X
		430	19.6	28.0	4.7	1-F08b	X				X		
		1152	8.9	9.6	15.7	1-F01		X	X				
		187	20.8	36.7	10.4	2-S05		X					X
		292	13.3	30.9	37.5	2-F05a		X					X
		286	19.4	37.0	12.4	2-F10	X						X
		124	36.3	64.9	7.5	2-S10	X						X
		126	15.0	30.8	8.2	2-F05b		X					X
		232	12.7	13.2	40.0	2-F01		X	X				
		165	12.3	17.6	12.3	2-F03		X			X		
		317	12.8	19.2	6.4	3-F08	X				X		
		625	10.0	10.5	27.0	3-F06	X		X				
		956	6.0	12.0	12.3	3-F03		X			X		
		466	8.9	24.8	4.5	3-F05		X					X
		178	9.7	16.0	4.3	3-F08	X				X		
		160	17.6	39.9	5.6	4-S05		X					X
		438	22.7	23.6	38.6	4-S06	X		X				
		792	13.8	15.4	15.4	4-S01		X	X				
		173	17.3	36.2	5.1	5-S05		X					X
		308	24.6	33.2	13.3	5-S08	X				X		
		69	21.3	40.2	12.6	5-S10	X						X
		195	13.8	28.9	9.2	5-S05		X					X
		198	38.7	53.4	1.0	6-F10	X						X
		340	29.7	30.4	0.0	6-F07	X			X			
		416	27.2	27.8	0.0	6-F06	X		X				
		308	35.2	36.3	13.3	6-S06	X		X				
		419	19.0	26.1	.6	6-S04		X				X	
		121	48.2	58.3	2.2	6-S05		X					X
		267	31.7	41.8	.4	6-F09	X					X	
		530	15.0	19.9	4.0	6-F03		X			X		
		220	37.7	49.9	8.4	6-S09	X					X	
		408	29.9	30.6	14.5	6-S01		X	X				
		134	22.1	44.4	7.2	7-S05		X					X
		168	38.2	67.1	5.0	7-S10	X						X
		332	20.0	32.4	4.8	7-S03a		X			X		
		346	28.3	31.5	1.8	7-S03b		X			X		
		295	33.2	33.6	0.0	7-S01		X	X				
		122	34.1	54.2	3.3	8-F10	X						X
		447	19.5	24.7	10.4	8-F03		X			X		
		322	26.5	33.6	7.0	8-F08	X				X		
		406	26.6	27.3	0.0	8-F06	X		X				
		364	14.2	29.0	5.6	8-F05		X					X
		330	8.3	15.5	8.5	9-F03		X			X		
		505	14.3	20.5	11.6	9-S03		X			X		
		1396	7.2	7.7	18.3	9-F01		X	X				
		485	8.6	22.0	7.1	9-F05		X					X
		648	9.4	17.2	4.8	9-F04		X				X	
		1675	6.4	7.2	15.9	9-F02		X		X			

Table D-1b. Data from the Proofreading Experiments

...ect	Plotting Symbol	Number of Lines Proofread	Average time (sec) for Zero Error Lines, T_z	Average Time (sec) Over All Lines, T_L	Percent Residual Error	File Name	Imaging Mode		Error Density				
							STS	VODEM	L1	L2	L3	L4	L5
		243	17.5	22.6	3.8	A-F03		X			X		
		425	21.9	28.1	9.2	A-S03		X			X		
		622	8.4	20.1	3.5	A-F05		X					X
		1430	8.0	8.3	14.8	A-F01		X	X				
		786	6.7	14.4	1.5	A-F04		X				X	
		1800	5.7	6.3	5.1	A-F02		X		X			
		150	14.8	36.0	14.8	B-F08	X				X		
		200	36.4	51.5	27.2	B-S08	X				X		
		491	10.2	21.3	7.3	B-F03		X			X		
		346	18.2	28.7	17.4	B-S03		X			X		
		113	10.0	26.3	7.0	B-F04		X				X	
		304	9.1	9.7	18.2	B-F01		X	X				
		81	11.6	38.2	2.2	B-F05		X					X
		228	6.8	7.9	5.7	B-F02		X		X			
	△	164	39.4	58.3	6.9	C-T03		X			X		
		548	21.9	22.6	15.3	C-S01		X	X				
		332	35.0	35.7	47.1	C-S06	X		X				
		204	52.1	60.7	39.4	C-S18	X				X		
		233	29.4	41.2	18.3	C-T03		X			X		
	◇	171	34.3	47.2	1.5	D-T03		X			X		
		699	16.2	17.1	9.1	D-S01		X	X				
		377	29.2	31.1	11.1	D-S06	X		X				
		304	30.6	41.5	2.0	D-S18	X				X		
		276	25.0	34.5	3.7	D-T03		X			X		
	□	266	18.8	30.8	11.3	E-T03		X			X		
		1227	9.5	10.0	60.5	E-S01		X	X				
		349	31.9	32.9	44.1	E-S06	X		X				
		277	32.6	44.6	27.7	E-S18	X				X		
		456	11.4	21.1	20.3	E-T03		X			X		
	○	240	27.5	33.7	.0	F-T03		X			X		
		651	18.5	19.2	7.1	F-S01		X	X				
		317	37.5	38.4	28.1	F-S06	X		X				
		342	32.6	40.0	15.5	F-S18	X				X		
		504	14.9	20.0	5.4	F-T03		X			X		

ORIGINAL PAGE IS
OF POOR QUALITY

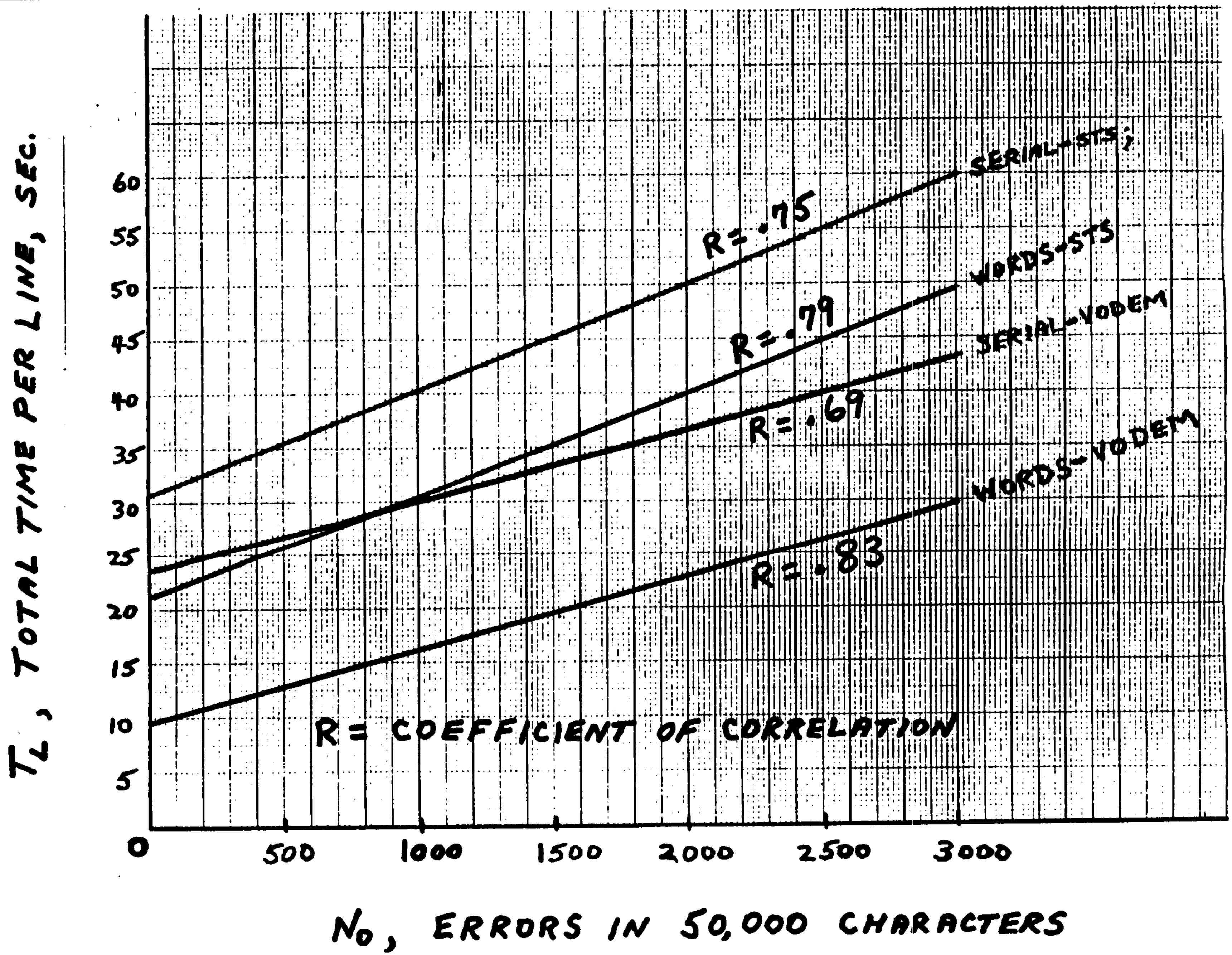


Figure D-1. Total Time per Line for Study A

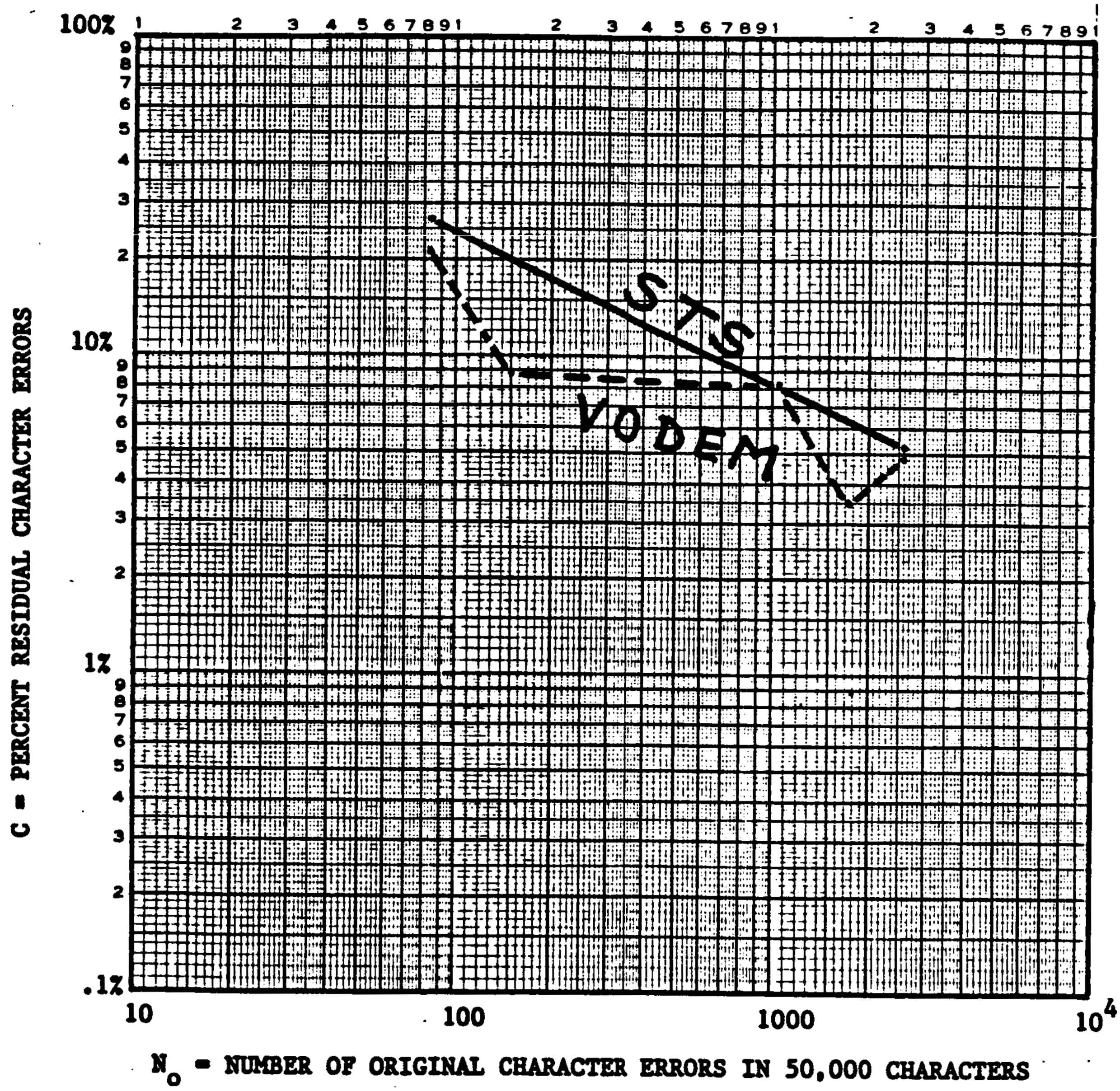


Figure D-2. Group Average Performance Charts for Study A, Words.

ORIGINAL PAGE IS
OF POOR QUALITY

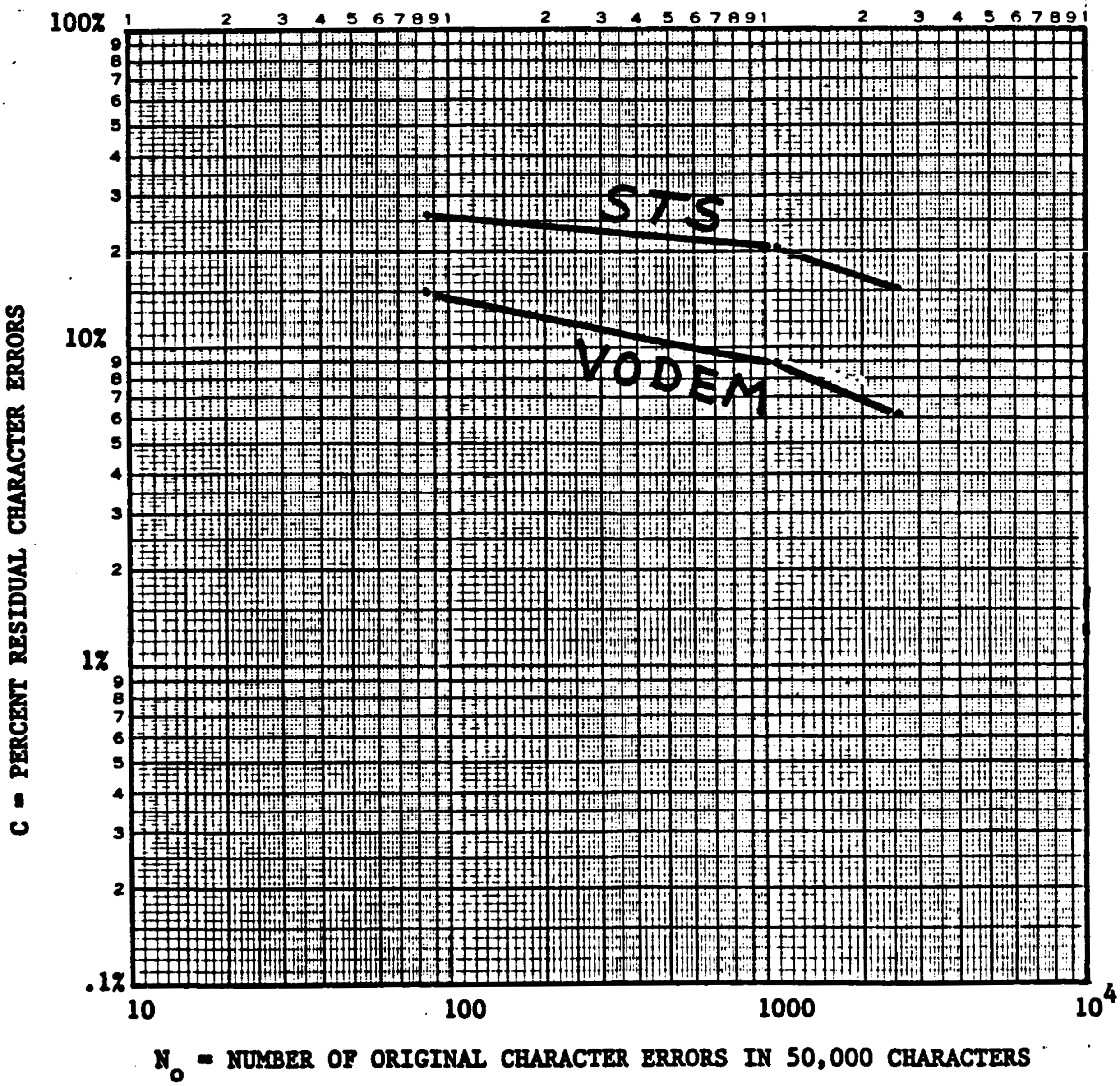


Figure D-3. Group Average Performance Charts for Study A, Serial Numbers

ORIGINAL PAGE IS
OF POOR QUALITY

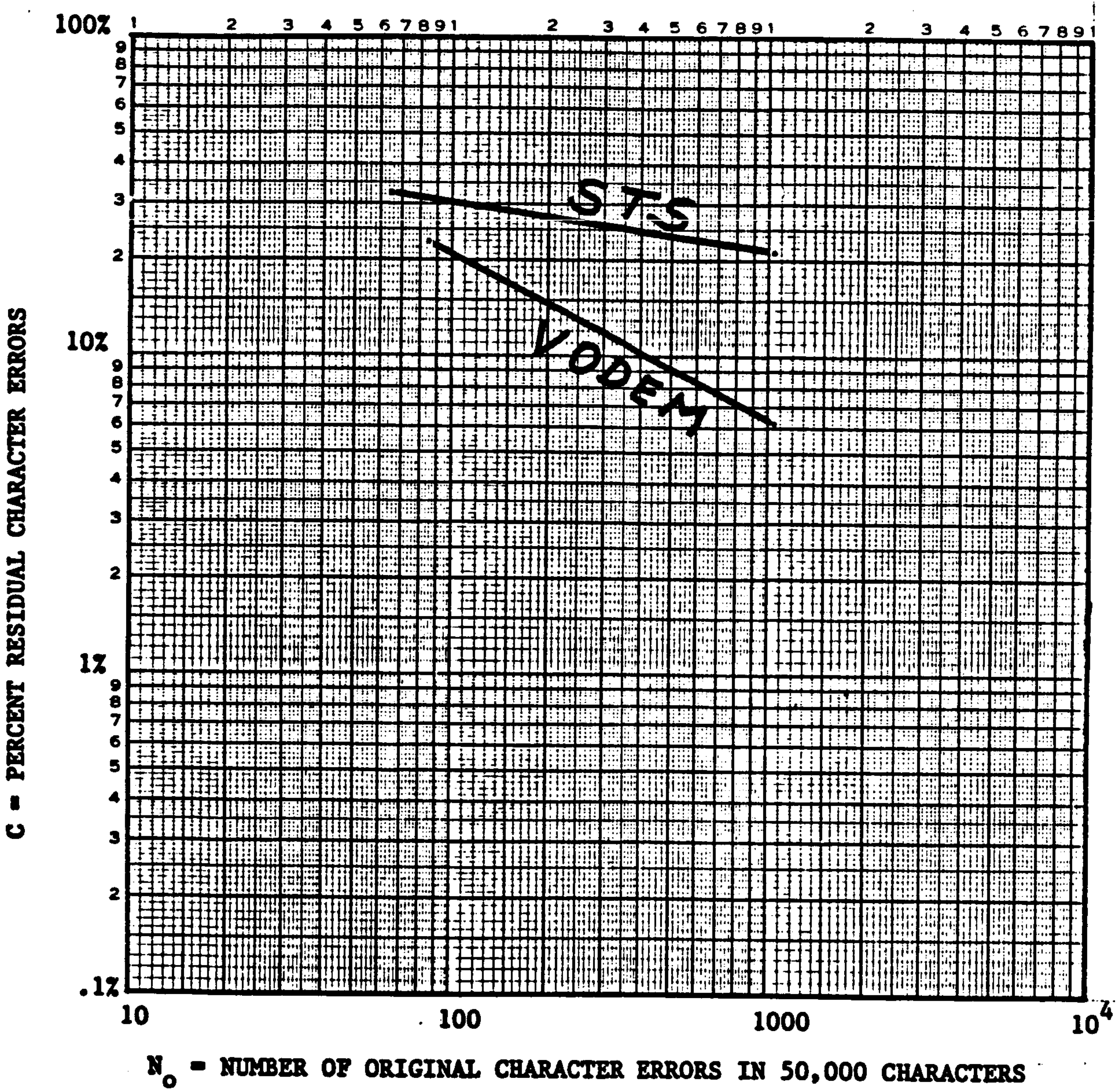


Figure D-4. Average Performance Chart, Study B, Serial

ORIGINAL PAGE IS
OF POOR QUALITY

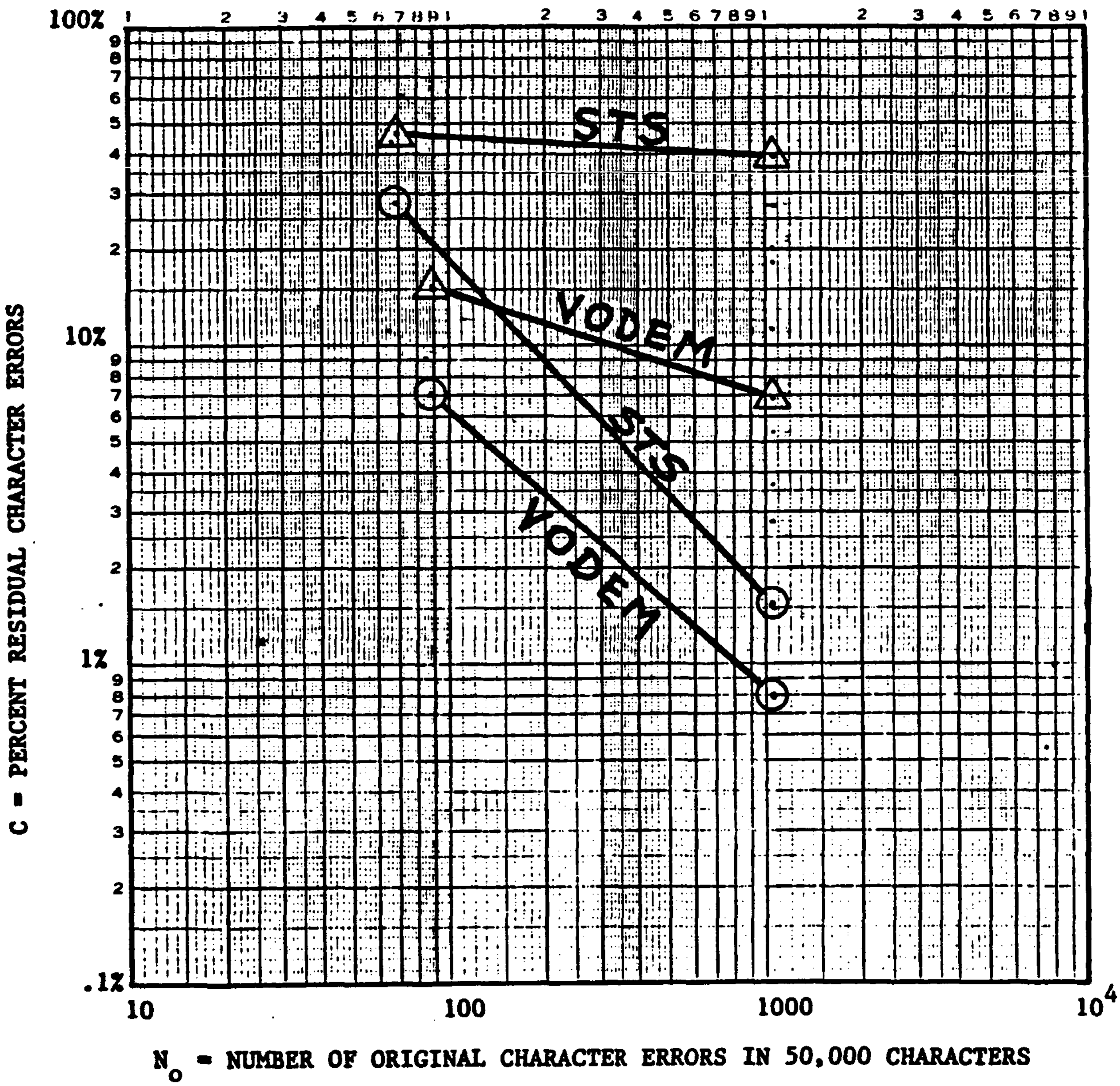


Figure D-5. Performance Charts for Selected Subjects from Study B

Section E

STATISTICAL EVALUATION

E.1 Introduction

The statistical evaluation of proofreading to eliminate error is quite complex. Numerous studies reported in the literature have been primarily directed toward achieving a lower error rate with little understanding of the physical, statistical, and mental processes involved. Almost all studies have had short test times and small test populations. One of the major problems in statistical evaluation is the availability of suitable test personnel with sufficient time. Test planning and duration as well as final results are affected. The normal amount of error resulting from keypunching and proofreading in testing and evaluation is related to the amount of training and experience each person has in the methods to be used. Some contractors, NASA and MICOM personnel, who volunteered for the test series, also had priority job responsibilities that did not permit them to complete the planned test time.

The predictive analysis for this program was initiated by an analysis of the work performed on an earlier Gamma Research proofreading program. The predictive statistical analysis was then continued through the first phase of this program that was performed by a group of government and contract employees (Study A) who volunteered to perform the proofreading tasks.

Time constraints prevented a similar analysis of the last group of personnel tested (Study B). Gamma Research appreciates the efforts of both groups.

E.2 Description of Statistical Program

A brief review of the variables involved in test error correction indicates the complexity of the problem. The variability and similarity among the symbols (characters) in the English language is indicative of the complexity of the problem. The appearance of errors in a string of symbols include: wrong characters, missing characters, added characters, wrong letter case, transposed characters and similarity in many letter geometries. Other errors which may appear include missing words and/or missing lines. In addition to the normal person-to-person differences, varied educational backgrounds and reading differences appear. Some personnel may have been taught to read and spell by "phonics" methods relying on phonetic features, others by word recognition. Persons taught speed reading are trained to look at a column of text at one time and consequently regress in single character recognition.

The mode of proofreading also must be considered. Computer data is commonly verified (proofed) by a second key input pass (and computer comparison). Book publishers and others have one person read copy to another. Side-to-side (STS) comparison of the old (dead or PDP(1)) copy with the new (live or PDP(2)) copy by typists and others is also prevalent.

A statistical approach for a major project must consider the objectives. Any that cannot be studied in detail should be randomly distributed so any effect the remaining variables may have will be as equally distributed throughout the experiment(s) as possible. The effect of unknown variables will be felt in all experimental work and is vital to planned statistical experimentation in which more than one variable is changed at a time. Multivariate analysis plans are therefore unsuitable for proofreading experiments (except those with very limited objectives) until more of the variables are fully understood. One of the aims of this study was to seek a method of controlling individual performance and, if practical, to find a general method of control that would aid in finding the cause of errors and improving individual performance. Engineering and psychological methods were applied to research the effect of error density and to determine the effect of imaging limited live and dead copy (PDP(2) and PDP(1), respectively) on the visual cortex of the brain simultaneously and so avoid errors due to recall from short term memory. One prime side effect of using a computer for the latter study is that it would permit objective measurement of the effect of accuracy and time. Computer programs designed for this study were written to fulfill these objectives. The project was limited to single unrelated words and serial numbers as test media. The RED computer program (see Section C) constructed word files for proofreading from a computer dictionary. The serial numbers generated by RED included familiar sets such as telephone, auto license numbers and automobile identification numbers, etc. After the dead (PDP(1)) copy was prepared, programmed errors were introduced to the live (PDP(2)) copy. The RED computer program randomly selected the lines and the position of an error in a word or serial number. The program imposed certain constraints on the characters. Where previous studies in the literature reported the frequency of a given error type the program approximated the frequency of these errors. To study the effect of error density five values of error density: L1, L2, L3, L4, and L5 were used. Two methods of proofreading were used. The normal side-to-side (STS) placement in which a computer printed dead (PDP(1)) copy was compared to a live (PDP(2)) copy on the computer screen. The computer correction program (PRF7)was used and a record made for analysis. The second method called the VODEM method placed a single line of the PDP(1) copy directly above the corresponding line of the PDP(2) copy on a CRT display and the errors caught by the proofreader were corrected by the same prompting method used for STS. Certain personal codes have been used in the statistical analysis to protect the identity of the test participants and to consolidate the report (see Section D).

The time available away from normal job responsibilities did not allow some people to complete the planned test program. In many instances the total number of samples used in the analysis do not meet the ideal statistical minimum. As anticipated, the lowest error density (L1 and L2) did not produce enough samples for even a practical statistical analysis. (For example, if only 20 samples are used, each one represents 5% of the data!) Therefore, the results in each individual instance must be used. The data used in this analysis is best represented by the data actually used as compared to the total test plan as shown in Figure E-1a and Figure E-1b.

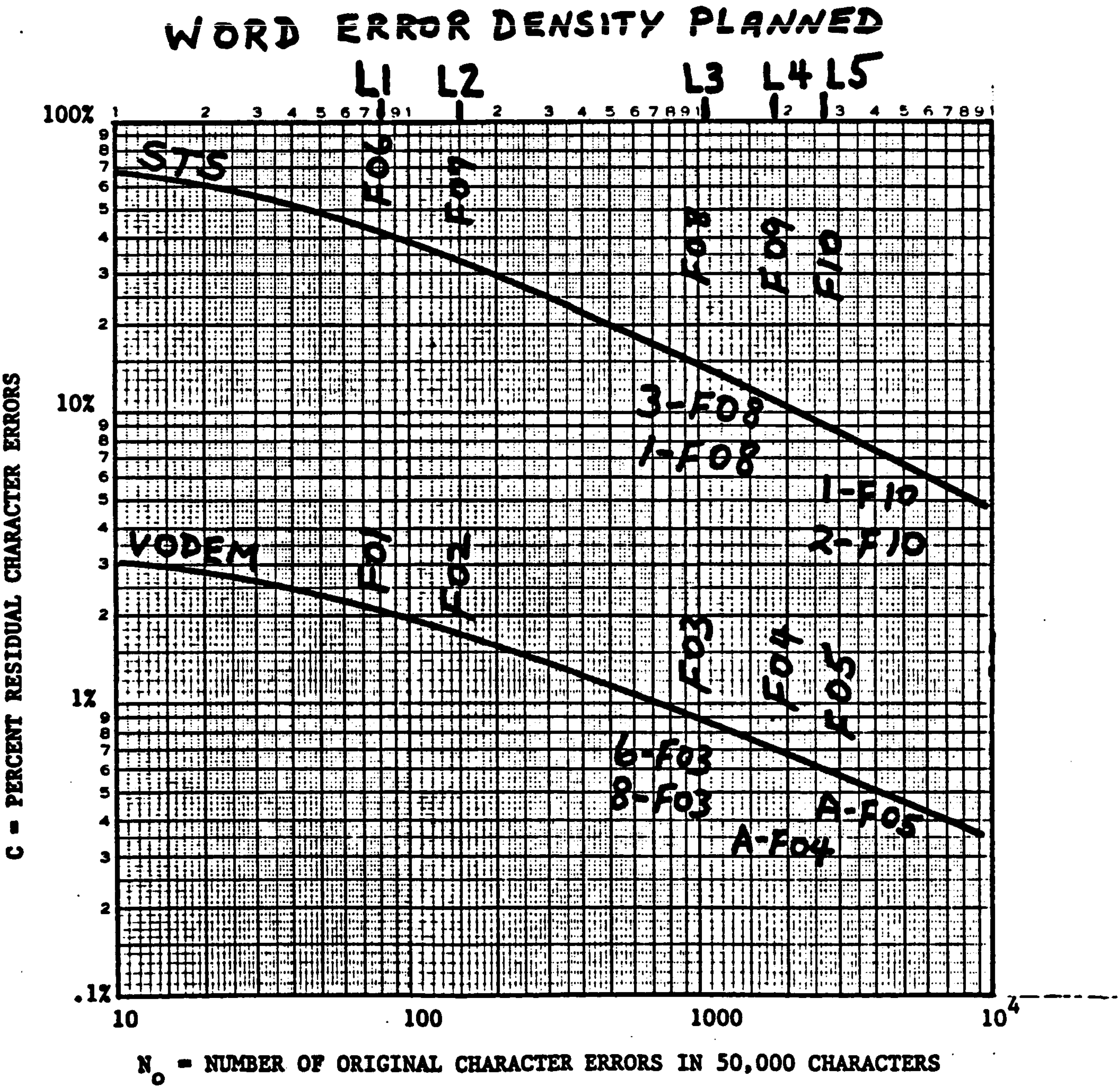


Figure E-1a. Data Used From Test Plan

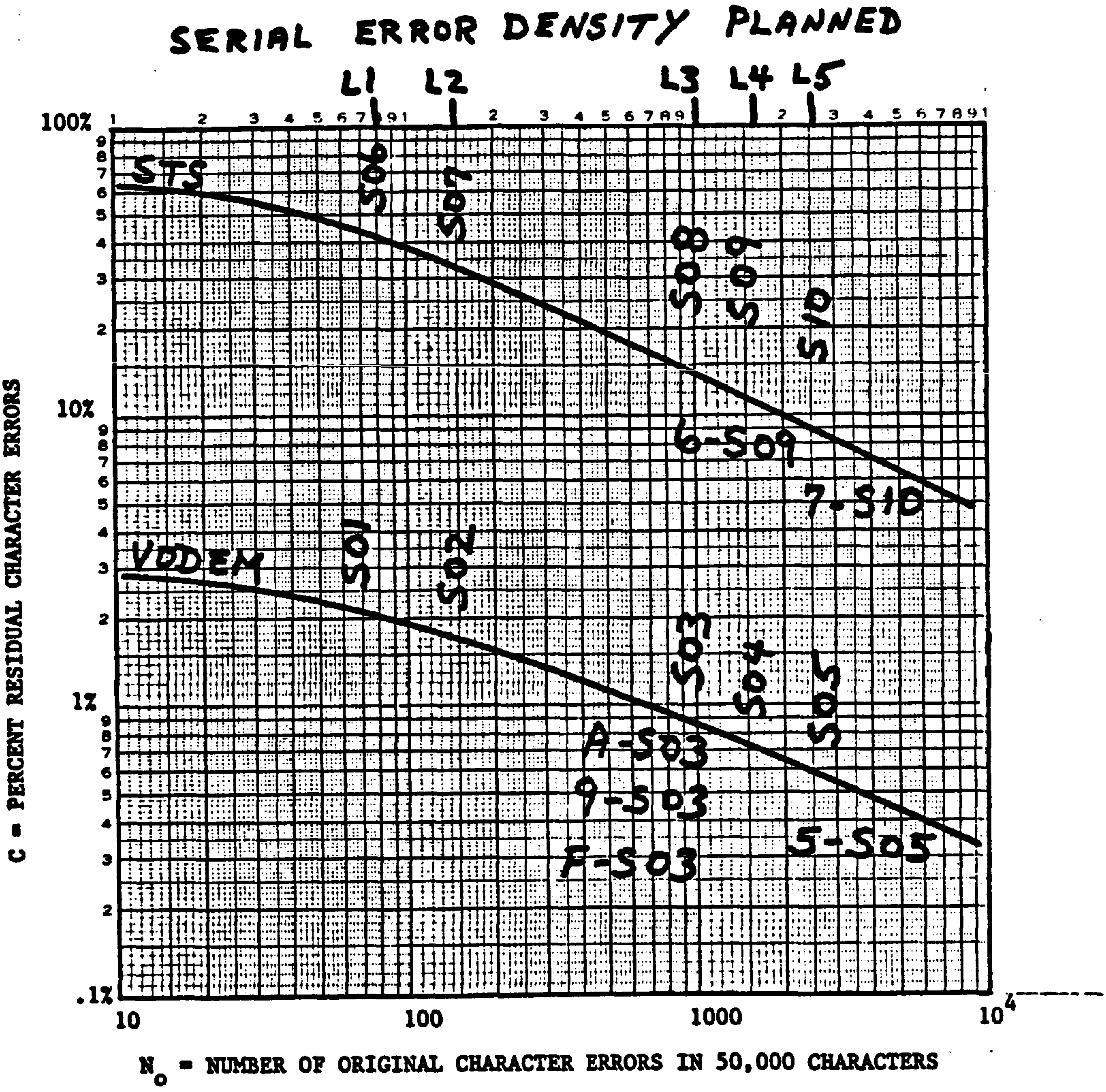


Figure E-1b. Data Used From Test Plan (Co...

Note: The curves for the STS and VODEM imaging modes are idealized examples in order to visualize the test plan. The error density values are defined in Section C.4.2. The File Names for each subject are defined in Table D-1. For example, A-S03 is a serial file at the middle error density (L3) for subject A.

E.2.1 A Brief Introduction to Statistical Analysis

Natural processes have been found to follow consistent statistical patterns when they are, in a general sense, free of constraints. When occurrences from a process are placed in an increasing order and the frequency with which each appears is plotted above each as a bar chart, the bars form the basis for a bell shaped curve. This curve is called a histogram. Figure E-2 shows a typical symmetrical bar chart histogram.

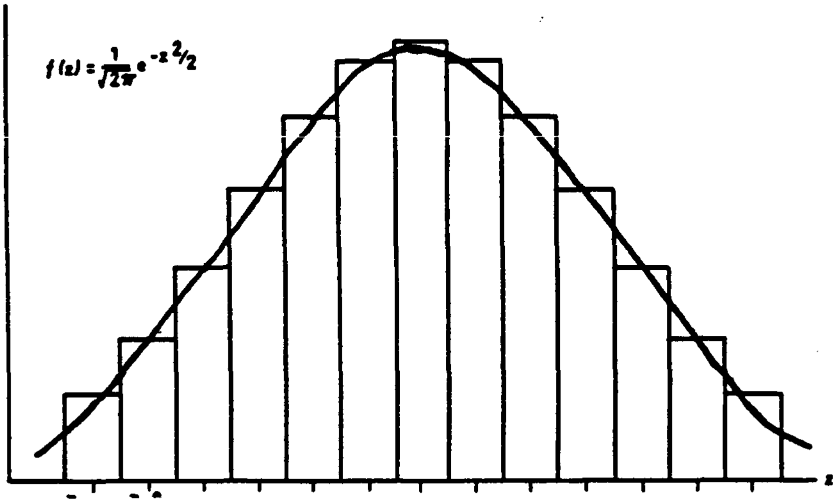


Figure E-2 Typical Bar Chart Histogram

The histogram found in most investigations is usually not so symmetrical but contains anomalies due to the chance of sampling or some constraint. When the bar chart histogram is smoothed into a curved figure a typical Gaussian or normal curve appears as shown in Figure E-3.

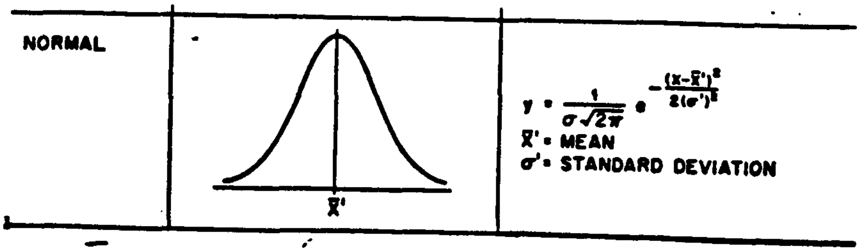


Figure E-3 Guaussian or Normal Curve

This curve is not only symmetrical about its average (mean), $\bar{x}$, but its median and mode coincide with the mean. However, in many instances one tail of the curve is constrained. For example, no chemical can be more than 100% pure nor can any be entirely free of impurities. The curve is then no longer symmetrical; in the first instance the average purity will move further toward the positive end of the curve and is said to be negatively skewed (Figure E-4a). If impurity is measured, the average or mean is moved far to the left and is positively skewed (Figure E-4b).

Figure E-4 Skewed Distributions

Dr. W.H. Shewhart who was responsible for the application of statistics to quality control in industry developed the Control Chart. This chart places the value of each event to be recorded in succession above the x or horizontal axis on a suitable scale. Control is now evaluated by the position of the values in relation to the values of the histogram whose base becomes the y or vertical axis of the control chart. These charts are greatly simplified by use of the standard deviation. The standard deviation (σ) is the square root of the sum of the individual differences from the mean (μ) and is symmetrical about the mean of a normal distribution. The mean plus and minus one standard deviation ($\mu \pm \sigma$) will include about two-thirds of all occurrences, ($\mu \pm 2\sigma$) about 94.5%, and ($\mu \pm 3\sigma$) about 99.7%. One additional property that is very useful for control is the fact that if the average of several occurrences (usually 4 or 5) is taken and plotted, the averages will tend to produce a normal distribution and control chart. Such a chart is called an x-bar chart ($\bar{x}$ chart). The $\bar{x}$ chart can now be interpreted (after checking for assurance that the distribution approximates the normal distribution) and simplified as the product of a normal distribution. The usual $\bar{x}$ chart is simplified by using a grid on which the mean and the upper and lower control limits (usually $\pm 3\sigma$) are highlighted. Figure E-5 illustrates the concept. Many other kinds of control charts can be used if interpreted in terms of their plotted distribution.

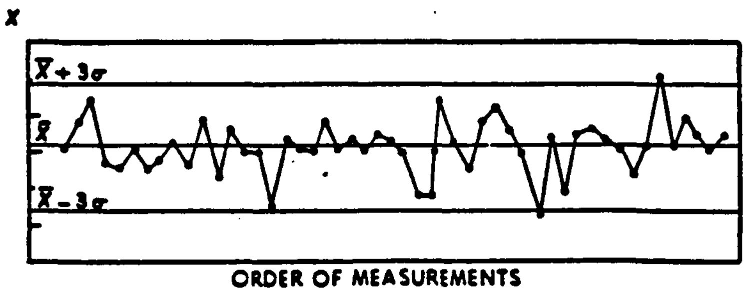


Figure E-5 Typical Normal Control Chart

E.2.2 Background to Program Statistical Analysis

A previous study by Gamma Research (Reference 5) using similar methods to aid handicapped people was reviewed early in the project with the object of developing methods that would objectively evaluate performance in proofreading and enhance their performance as well. One of the better methods of providing such assistance and control is the statistical control chart (see Section E.2.1 for some elementary detail). The usual approach to this problem is to transform the data to a normal distribution chart. A "fraction defective" or P-Chart was used in the earlier program. The chart was based on the fraction not corrected in samples containing 20 errors in the live copy. This chart was useful in obtaining comparison of performance and personal progress but the data pattern (histogram) was not normal. The usual technique of averaging an increasing number of samples for plotting was tried to get a normal distribution of data as the initial step in getting a sound control technique. An increasing sample size was tried next and it was found that even samples of 100 originally erroneous characters did not yield a normal distribution. Reference was made to the statistical literature to determine the underlying frequency distribution of residual error (I_f) in various sample sizes of original errors (I_o) in the live copy. Techniques advocated by the 4th edition of Sir William Elderton's classic "Systems of Frequency Curves" (Reference 11) gave some insight but no definitive results. The data generated by some (but not all) personnel in the earlier Gamma Research study using handicapped people were found to approximate an exponential distribution (Figure E-6).

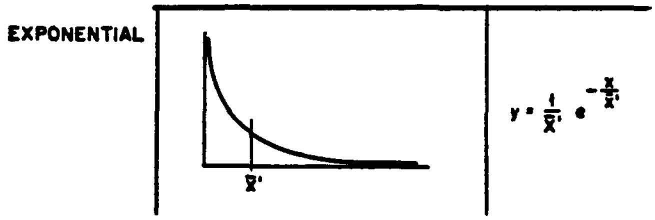


Figure E-6 Exponential Distribution

The form of this curve led the study toward investigating Gumbel's extreme value theory and distribution. Details of this theory are not found in the usual statistical texts. A copy of Gumbel's original paper (Reference 12) and his probability paper were found with Mr. Don Royston of Morton Thiokol, Huntsville. Mr. Royston offered further assistance. He used Gumbel's extreme probability paper in his analysis. Probability paper is designed with one axis of a graph spaced so the cumulative probability being considered produces a straight line when plotted against a linear or derived function of the parameter being studied. Conversely, the parameter studied may be considered as having the frequency distribution of the probability paper used. Figure E-7 shows Gumbel's probability paper and a typical plot.

Data taken during an earlier Gamma Research study (Reference 5) was used. The data was from proofreading tests of representative handicapped personnel with 2 weeks average training in Gamma Research methods. Figure E-8 shows Gumbel's Distribution for trained personnel. Some of these people had never used a computer before.

The linearity of the plots for individual results and their average is a strong indication that the Gumbel Extreme Value theory applies to these people and will apply to most future personnel tested. Because Gumbel's basic work is not readily available some details are given here.

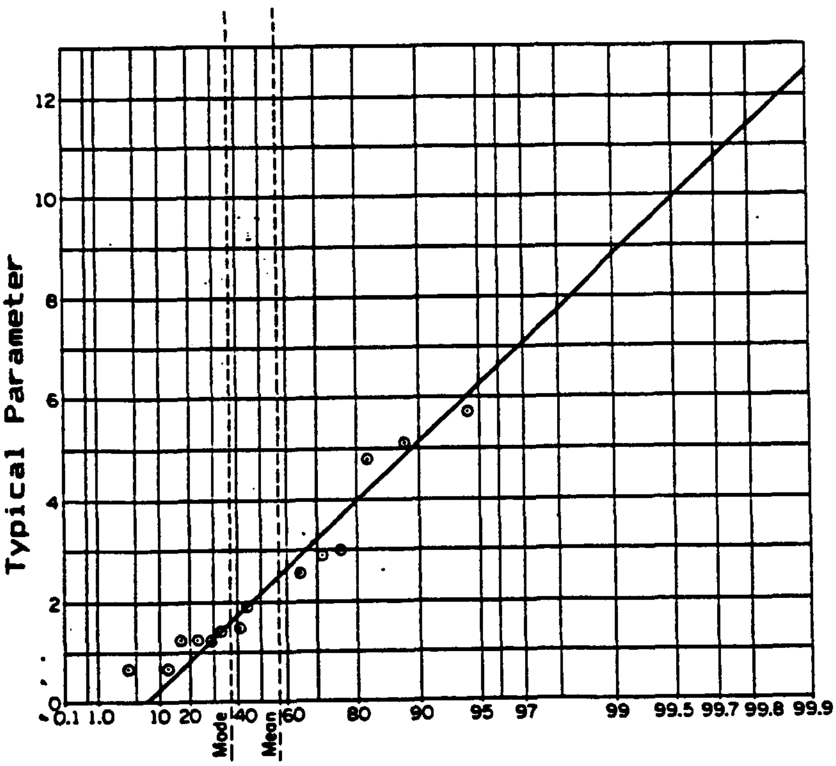


Figure E-7 Typical Gumbel Extreme Probability Plot

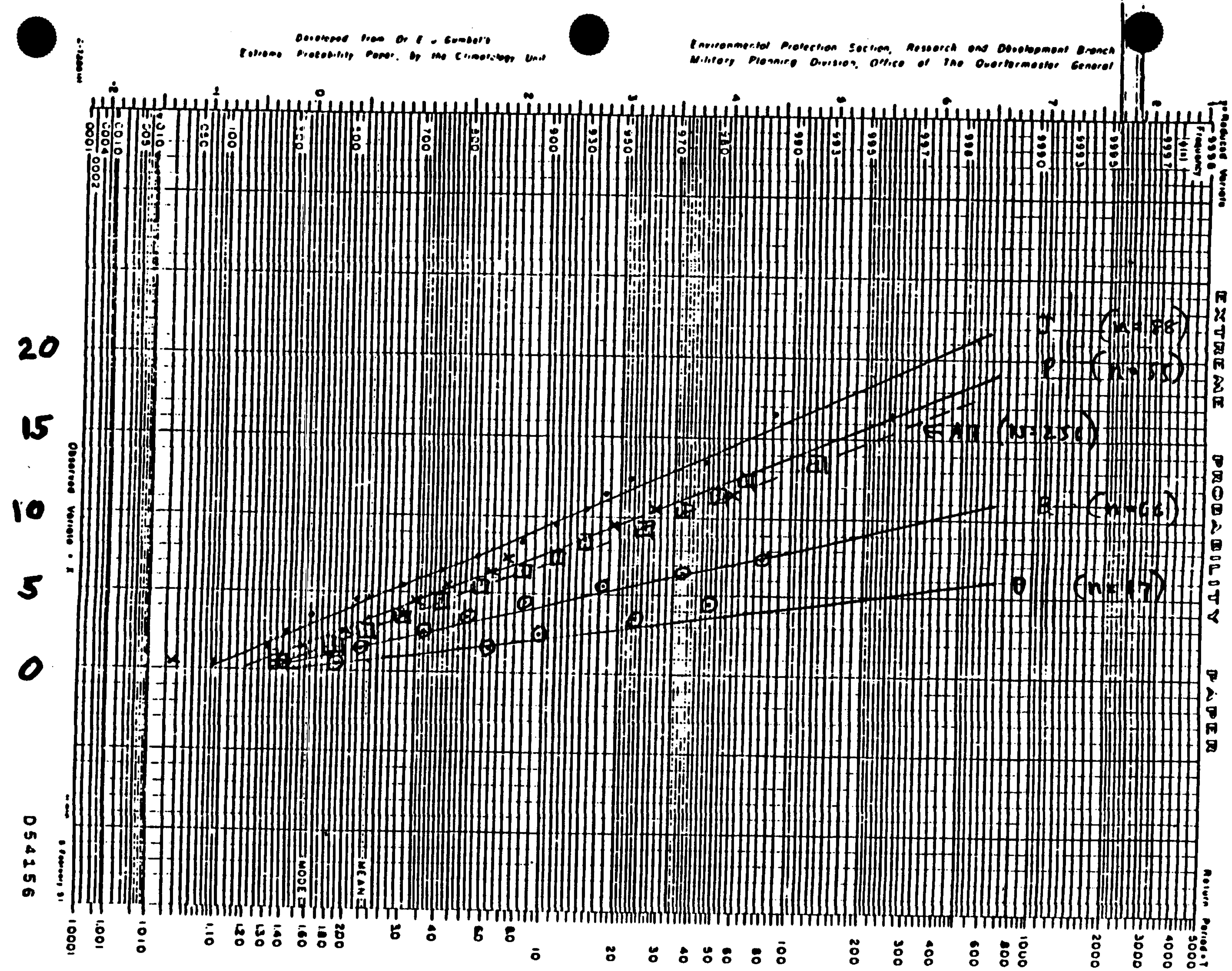


Figure E-8 Gumbel's Distribution for Trained Personnel

E.3 The Gumbel Distribution

While searching for a way to predict floods E.J. Gumbel of the National Bureau of Standards found that the annual maxima of stream flow formed a distribution of their own that could be used to predict the return period of floods of a given height better than any other statistic. Shortly afterwards he and others found his distribution could predict both the maxima and the minima of not only floods but also of other phenomena as well. Some examples are: capacitor breakdown voltages, wind-gust maxima, human longevity, time interval for the next emission from radioactive materials, draught periods, extinction of bacteria with disinfectant exposure, meteorological pressure extremes, and the life of light bulbs. Gumbel developed a probability paper for the distribution. Probability paper gives a simple graphic method for testing the fit between observation and theory, avoids laborious calculation. It may be used as the basis for determining process levels. Let x be a continuous variable that will usually have a dimension in cycles, time, length, and as in our case residual error as a function of an original error sample.

$$x = u + \frac{1}{\alpha} y$$

where u and $1/\alpha$ are parameters of x and y.

u is a selected average, e.g., mean, mode, median

$1/\alpha$ is a certain measure of dispersion

(See Section E.5 Numerical Analysis of Extreme Probability Data for more detail)

Probability paper is a rectangular grid on which x is plotted on one axis and the probability variate $(\Phi(Y) = FX)$ is plotted on the other axis.

The Gumbel probability paper has the observed (x) value plotted vertically and the reduced probability variate (y) plotted horizontally. This eases the utilization of the least squares method of plotting the regression line (when warranted). The chart is prepared by placing the observed values in ascending or descending order and the occurrence of each of the observed values determined. The respective cumulative values of the ocurrences are determined along with their summation. Each value is now divided by the total of all occurrences plus one. This procedure allows plotting all points and introduces an insignificant error when the size of the occurrences is considered in relation to the universe from which it was taken. Gumbel has developed several other scales related to the selected probability scales that may be obtained from his referenced paper (NBS Applied Science Series #33) that have not been discussed herein.

When the actual data and the probability paper have matching distributions, the data will plot as a straight line. Some mismatch outside fifteen or eighty-five percent is usually not significant. When straight lines plotted from the data intersect the data is skewed in relation to the distribution of the paper. Similarly, lines that do not intersect or only intersect through a curve indicate a bimodal relationship between the 2 parts of the data that are each following the distribution of the paper. In the latter case, histograms would overlap on adjacent tails.

The use of the Gumbel theory is appropriate to this proofreading study because the data are the residual errors left after proofreading each group of 100 errors embedded at random in the body of the word and serial number tests.

E.3.1 Preparation of Gumbel Plots with Probability Paper

A step by step method of preparing Gumbel Probability Paper plots follows for the reader's convenience:

1. Examine the data and reject values only for cause. For example: Computer summary would report about seventy-five errors if the operator did not note a missing line (incorporated in tests to indicate typing speed).
2. Tabulate the number of errors reported by the computer per 100 error samples in ascending order (or with any other sample size).
3. Calculate the ascending cumulative errors for each sample. This is usually done by a rough histogram using the tally count (卌) for each sample number that has errors. The base (x) line includes samples with no error.
4. Calculate the cumulative probability of error for each sample. By using the total error count +1, all points may be included in the plot with negligible error in relation to the total sample number as it relates to the total population universe.
5. Plot the cumulative probability versus all the possible error counts over its range.

Example:

Given an error count (the number of error left (I_f) after a proofreading from a sample containing I_o errors within the body of the proofreading material):

1. Given the example:

0	1	2	3	Number of errors in any sample
20	15	10	5	Total occurrence of the number of errors in each sample for the work performed

2. Calculate the summation over each number of occurrences:

$\sum$ = 20 + 15 + 10 + 5 =50 ($\sum$ = N, number of samples)

3. Divide by N+1 (i.e., 51) and accumulate cumulative probability of occurrence.

20 / 51	=	0.3922 :	cumulative sum is 0.3922
15 / 51	=	0.2941 :	cumulative sum is 0.6863
10 / 51	=	0.1960 :	cumulative sum is 0.8823
5 / 51	=	0.0980 :	cumulative sum is 0.9803

4. Plot the result on Probability Paper (some Probability Papers are scaled as percentages).

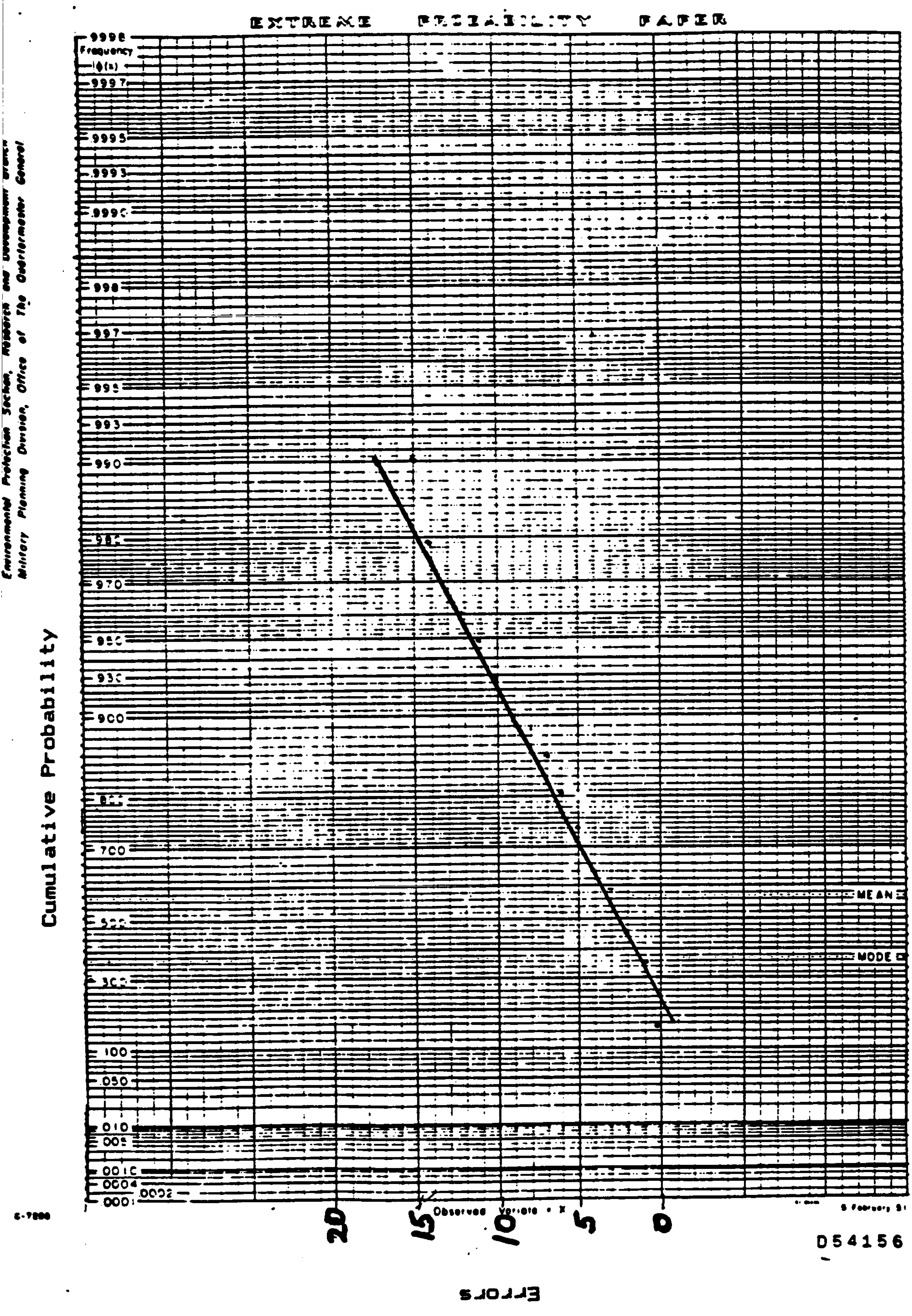
Note: This is an example only. Ideally there should be a minimum of 20 samples. The use of N+1 allows plotting all samples with negligible error in relation to the universe.

E.4 Test Results.

There was great variability in both the quality and quantity of each individual's proofreading results. Much of the variation may well be caused by the short time they were able to devote to the testing and lack of experience with the method(s).

The smaller samples allowed a greater variation in the comparison between the work of individuals. However, when all the work produced was considered as a single proofreading task, the summation regardless of content, the mode of performance, and even individual variation showed a remarkable agreement with Gumbel's Extreme Value Theory. Figure E-9 shows the overall indication that proofreading error follows Gumbel's extreme probability distribution despite the variation in people, test type and test mode.

A series of precontrol charts were prepared for this study along with the associated Gumbel charts. The precontrol charts are those made while accumulating data for "control" charts. The initial control charts are the second step after getting a knowledge of the underlying distribution, and may at times be used as an aid for determining it. It is very important that a user determine the distribution when data is delayed or expensive to get. If the distribution is known and probability paper is available the control chart may be used to get quantitative data about process performance. Figure E-10 shows how the probability axis of Gumbel probability paper may be used as a guide to get a quantitative scale for a control chart. The Gumbel probability paper shown in Figure E-10 was placed diagonally to match the total width of the vertical axis on the precontrol charts that will be shown later. For example, the expected average and mode, as well as their usual control chart limits, can be estimated. The reduced probability scale was placed on two of the word-vodem charts and the theoretical Gumbel mean and 3σ limits were marked to illustrate the point (see Figures 13 and 14a shown later in the text). There are a series of interpretations of control charts published that may not be used here because they are for normal distributions and are not true for the Gumbel extreme probability distribution. While both of the charts are within limits it must be remembered they are made from only a few samples and as sample sizes increase the chance of large values increases. It should be remembered that the steps taken here are only preliminary to true statistical process control charts. It is the author's experience that very few start-ups are in statistical control. Western Electric Company (Reference 13) authors in fact recommend at least three periods of process sampling to study the process capability of an ongoing process with a total of at least 100 samples and then point out that the statistical charts will probably be out of control. They also point out that only after many repeat analyses by competent engineers will most processes come under the statistical process control that assures an accurate product at the lowest cost.



D54156

Figure E-9 Summation of All Samples Used (words, serial numbers, STS, VODEM, error density)

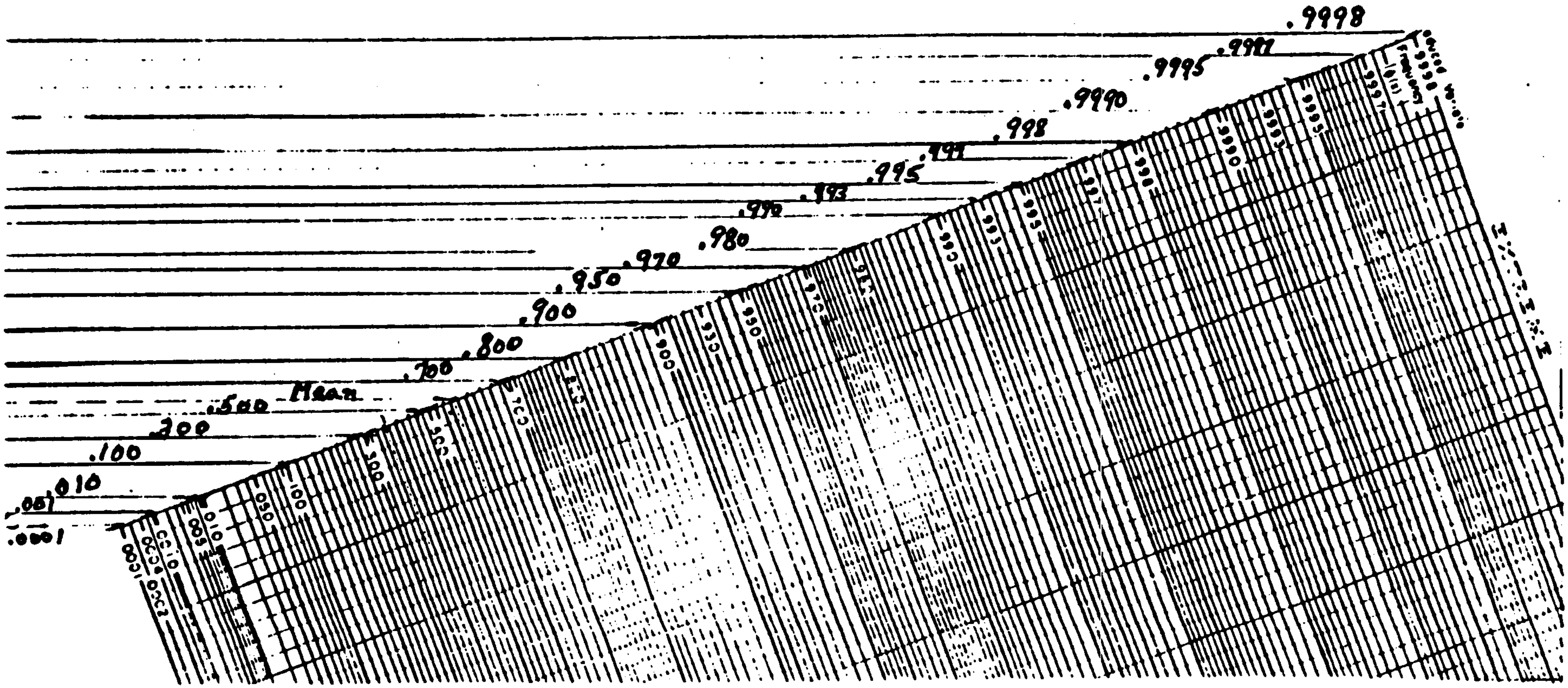


Figure E-10 Changing Probability Scales

Precontrol charts of each person's work are included primarily as examples. Where there is an obvious departure of an individual's work from Gumbel's distribution, the precontrol charts are examined to see if the departure may be related to the sequence of that person's performance. On-line reporting of such departure(s) in the future could aid in isolating and correcting the cause of such erratic performance.

E.4.1 Word Proofreading

A Gumbel probability graph (Figure E-11) was prepared for proofreading words by 4 people. The straight lines for three of the 4 had very similar slopes. The fourth person (A-F04) showed a bimodal distribution (in relation to Gumbel's distribution). The line representing one portion of this person's work was essentially parallel to those of the other three. Since the remaining straight line portion of this person's work was set by only 2 points, special attention was given to the corresponding precontrol chart (see Figure E-13b for person A-F04, shown later in the text). This chart has no apparent anomalies so the bimodal characteristic may be real or occurs by chance alone. This type of characteristic will be watched in future work and its cause investigated. An on-line summary would be particularly useful in this type of investigation.

Proofreading words by the side to side (STS) method shows considerable difference in the slopes of the extreme probability curves of the individuals (see Figure E-12). They indicate a higher error rate than the VODEM method and a wider deviation. One person (3-F08) on the medium density material has a curve that would indicate a skewed distribution. The corresponding control chart (Figure E-15b) is very erratic with a high initial error rate that drops rapidly but does not maintain its low level. Again this type of performance could be traced better with an on-line analysis.

E.4.2 Proofreading Serial Numbers

Proofreading serial numbers is more difficult. The familiarity with words and their spelling is apparently helpful in proofreading despite the various methods of teaching reading. Each character in a serial number must be read because the proofreader has no prior knowledge of a sequence. Errors still generally follow the Gumbel distribution.

Examining the Gumbel chart for proofreading serial numbers by VODEM (Figure E-17) shows an increased dispersion of the slope of the individual curves for words. One specific individual (5-S05) reading a high error density does not follow the Gumbel distribution. Examination of the precontrol chart (Figure E-20b) again offers no explanation and our manual analysis methods uncovered the information too late for effective personal interviews.

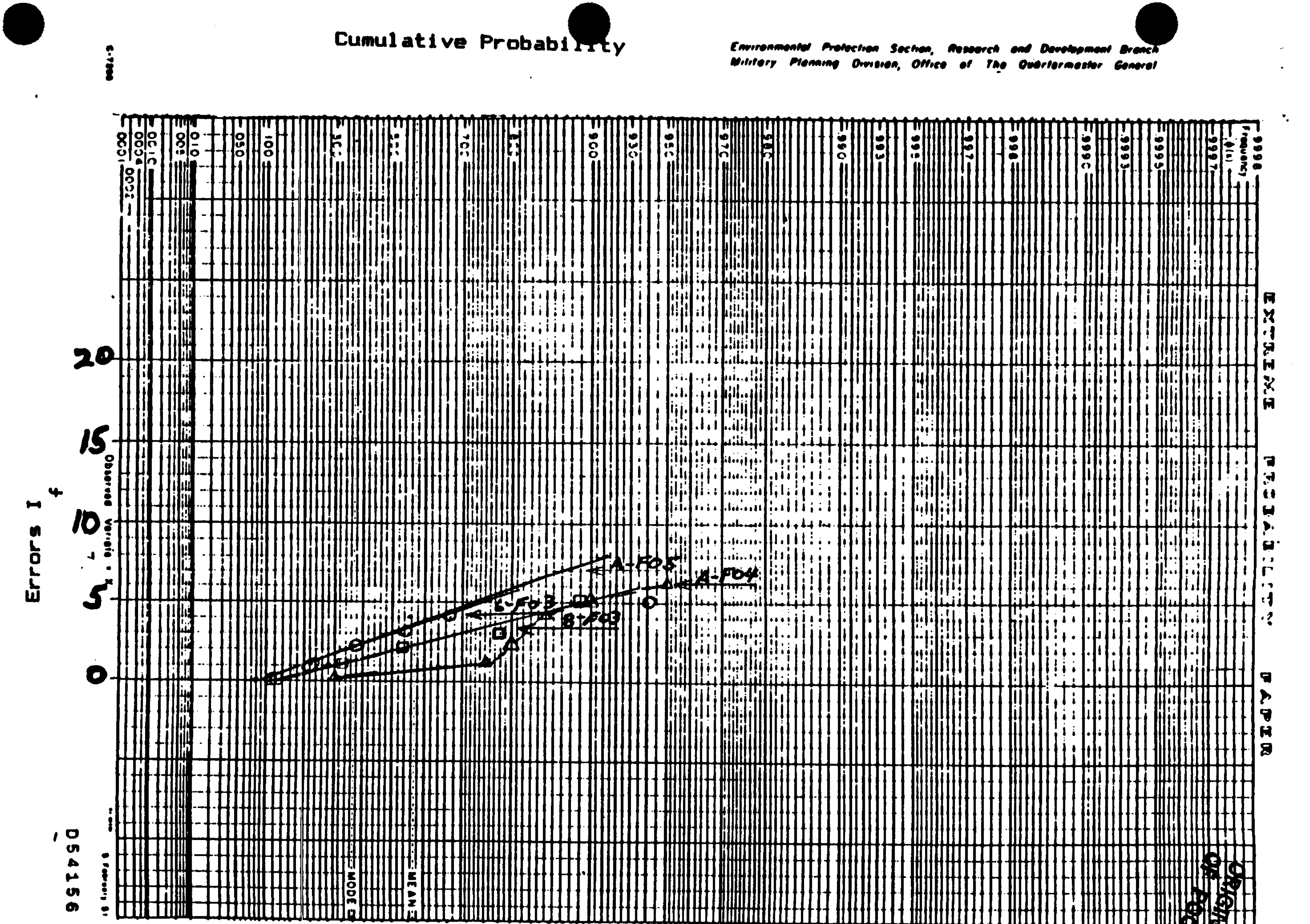


Figure E-11 Proofreading Results of Words Test by VODEM

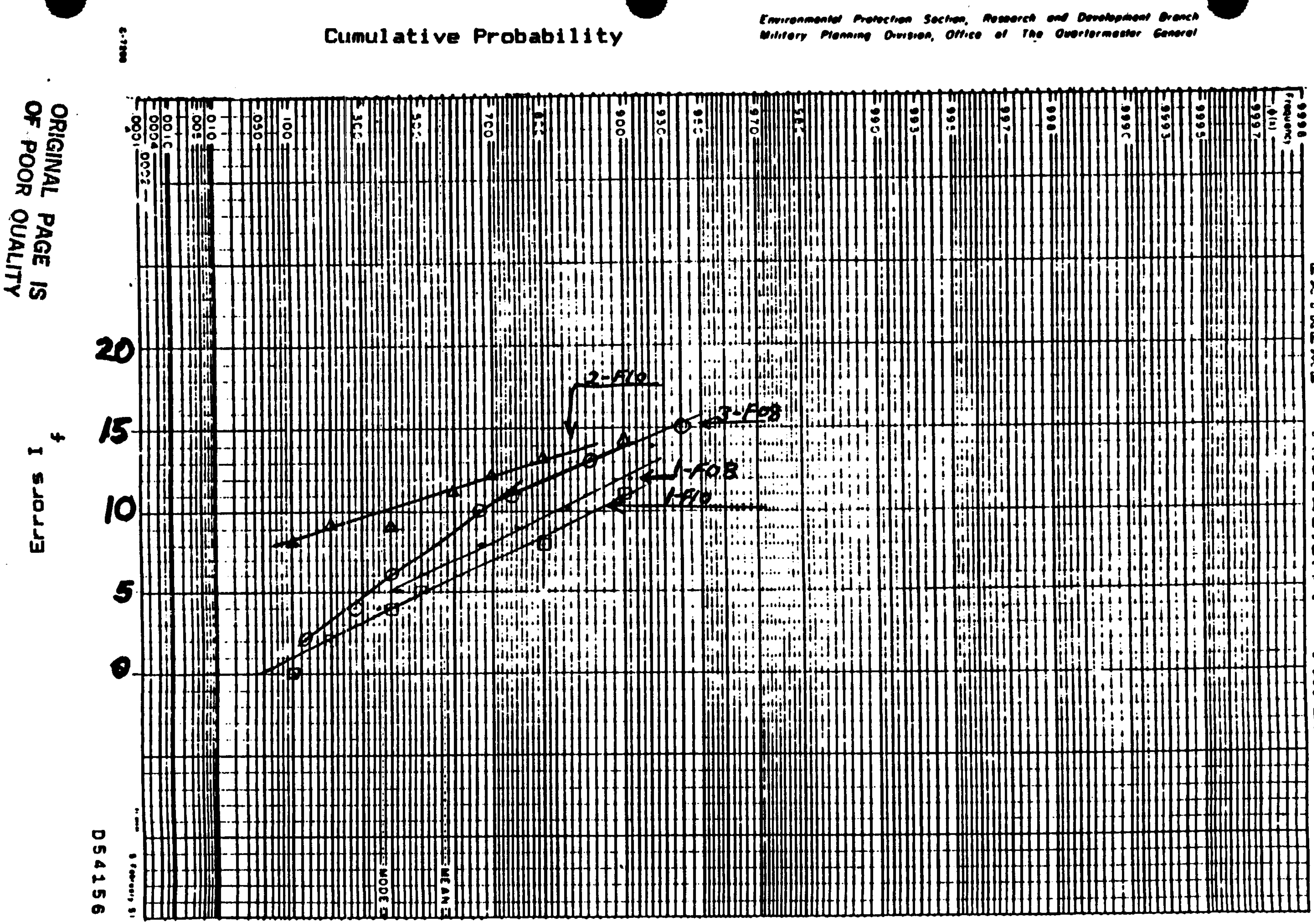


Figure E-12 Proofreading Results of Words Test by STS Mode

ORIGINAL PAGE IS
OF POOR QUALITY

Figure E-13 Individuals Precontrol Chart of Word Test by VODEM

Figure E-13a

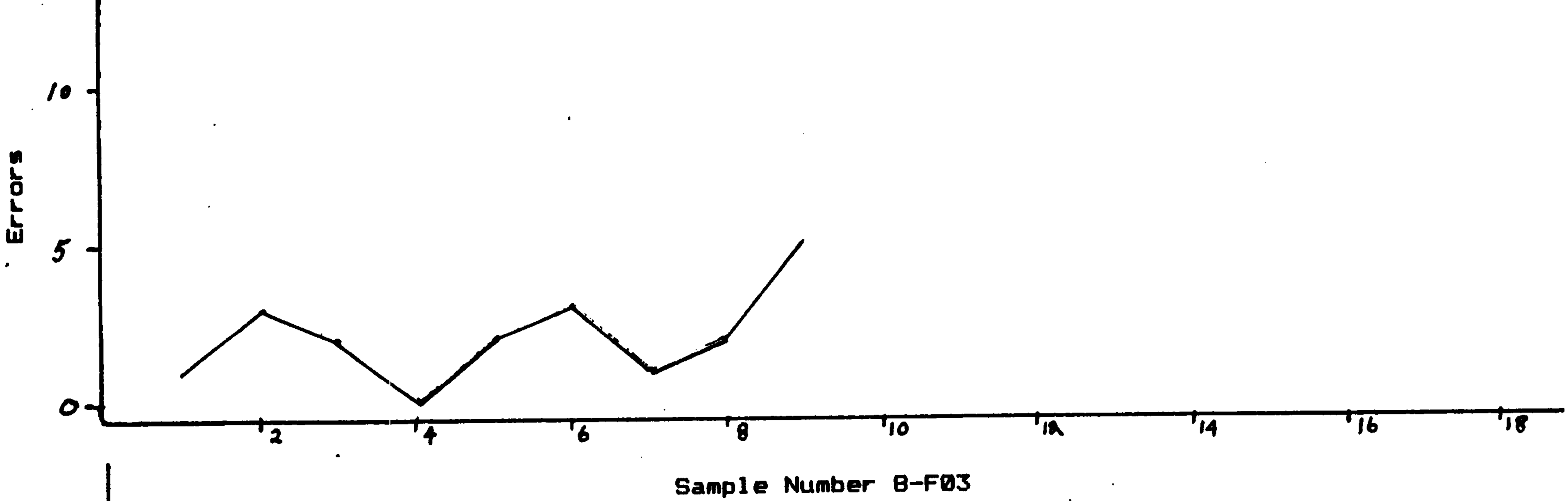


Figure E-13b

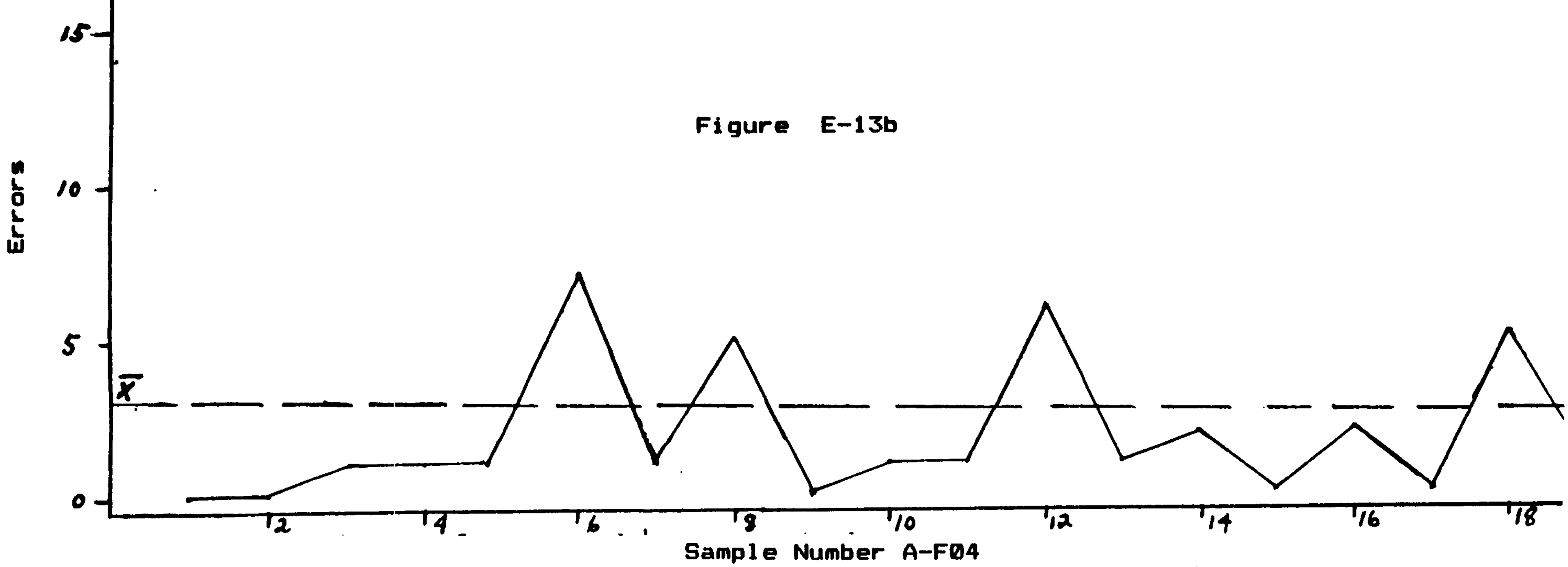


Figure E-14 Individuals Precontrol Chart of Word Test by VODEM (com)

Figure E-14a

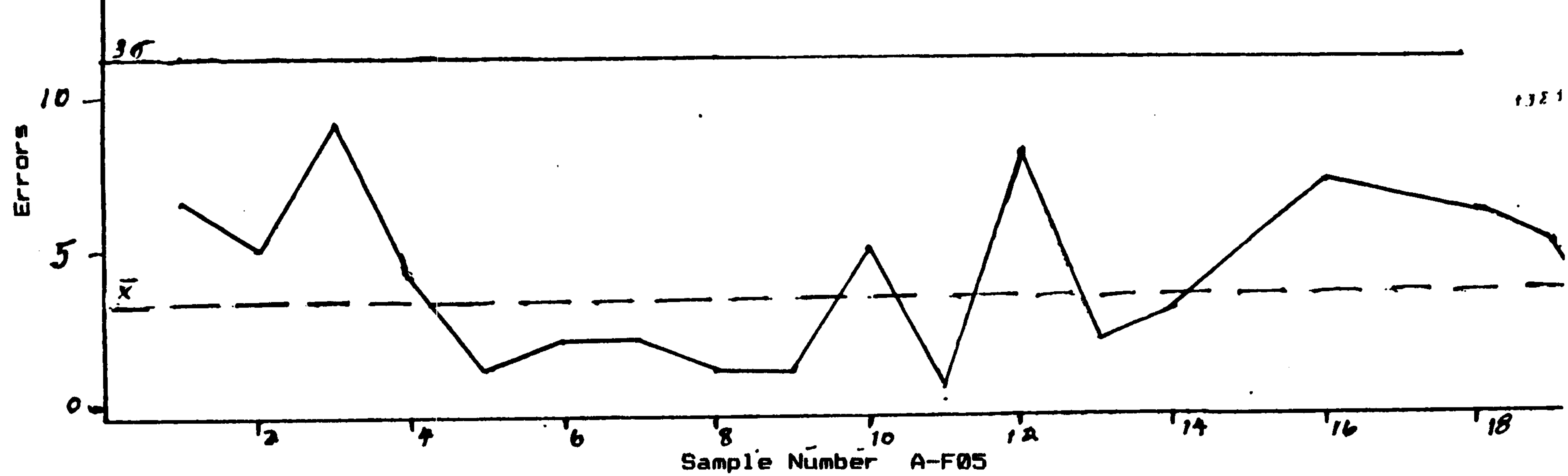
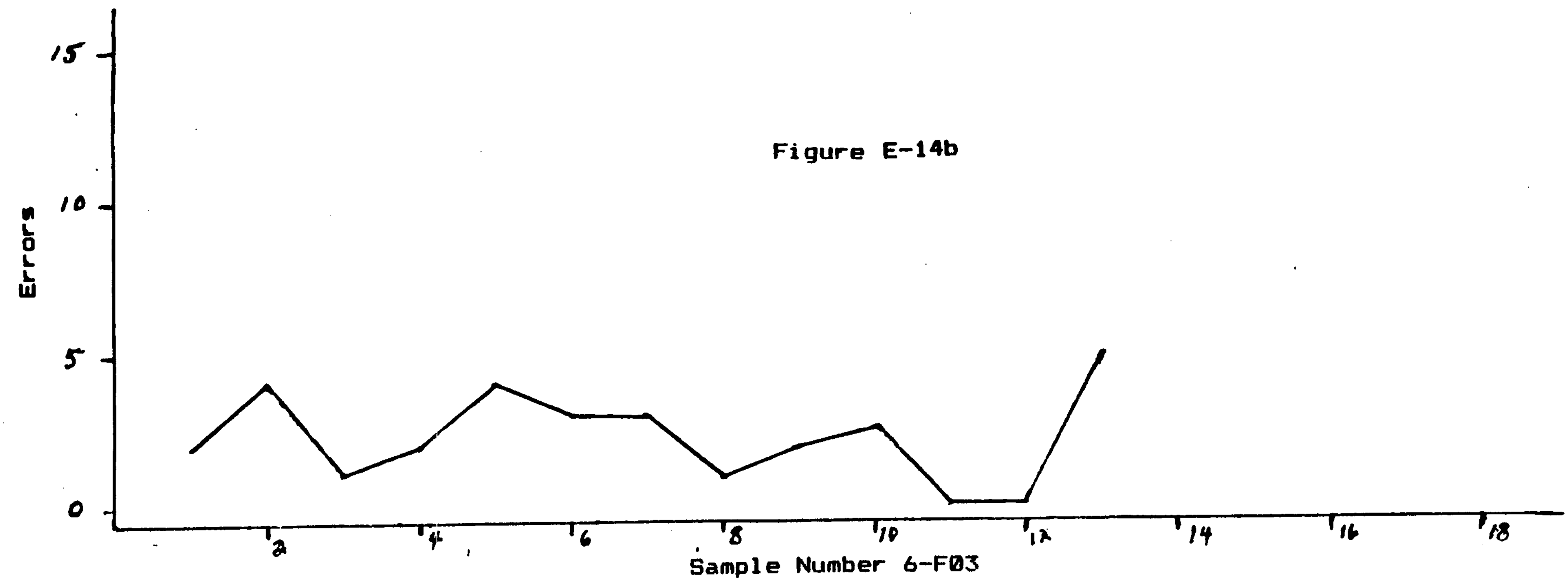


Figure E-14b



Individuals Precontrol Chart of Word Test by STS

Figure E-15a

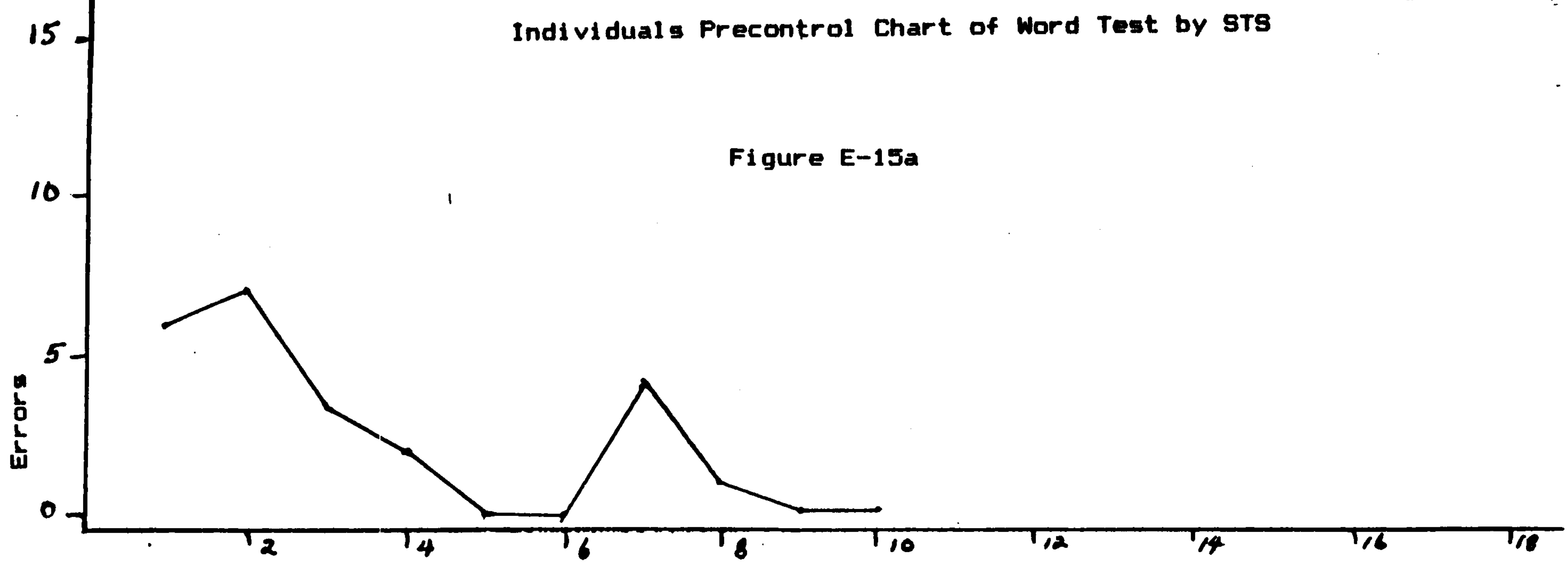
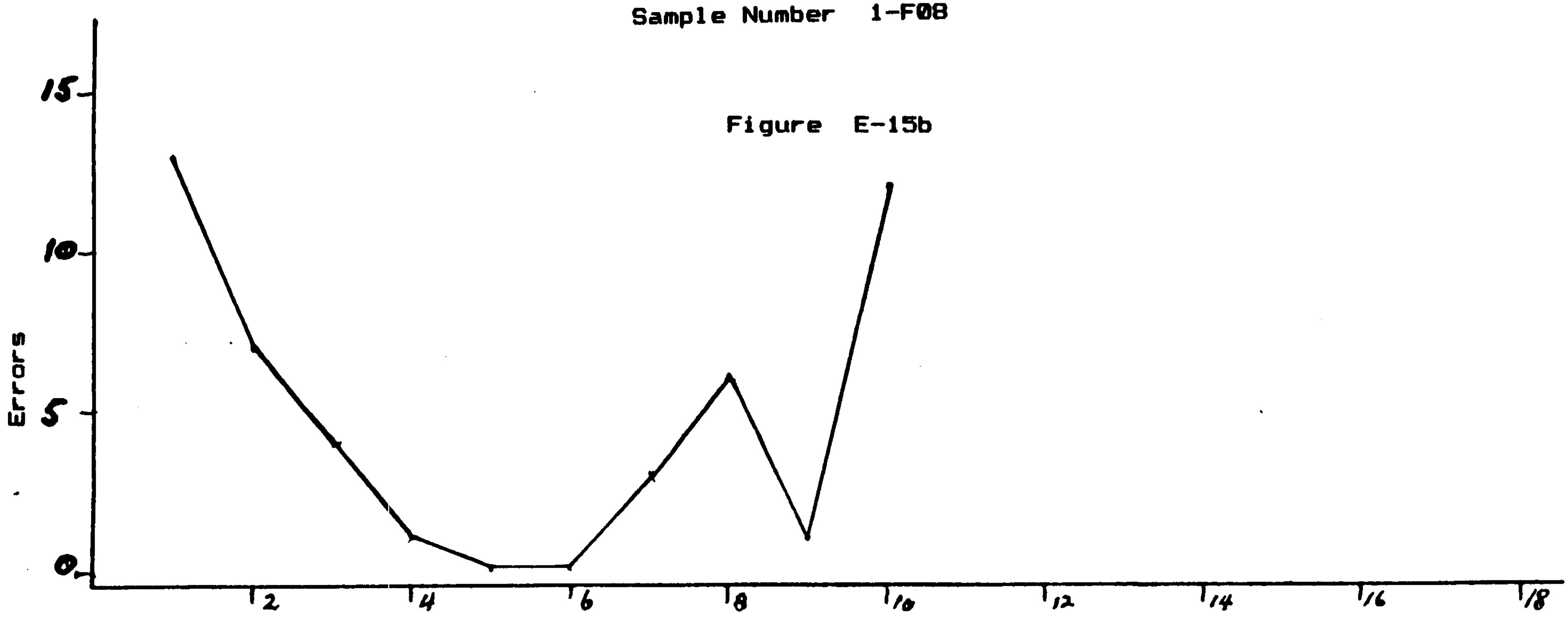


Figure E-15b



Individual Precontrol Chart of Word Test by STS (continued)

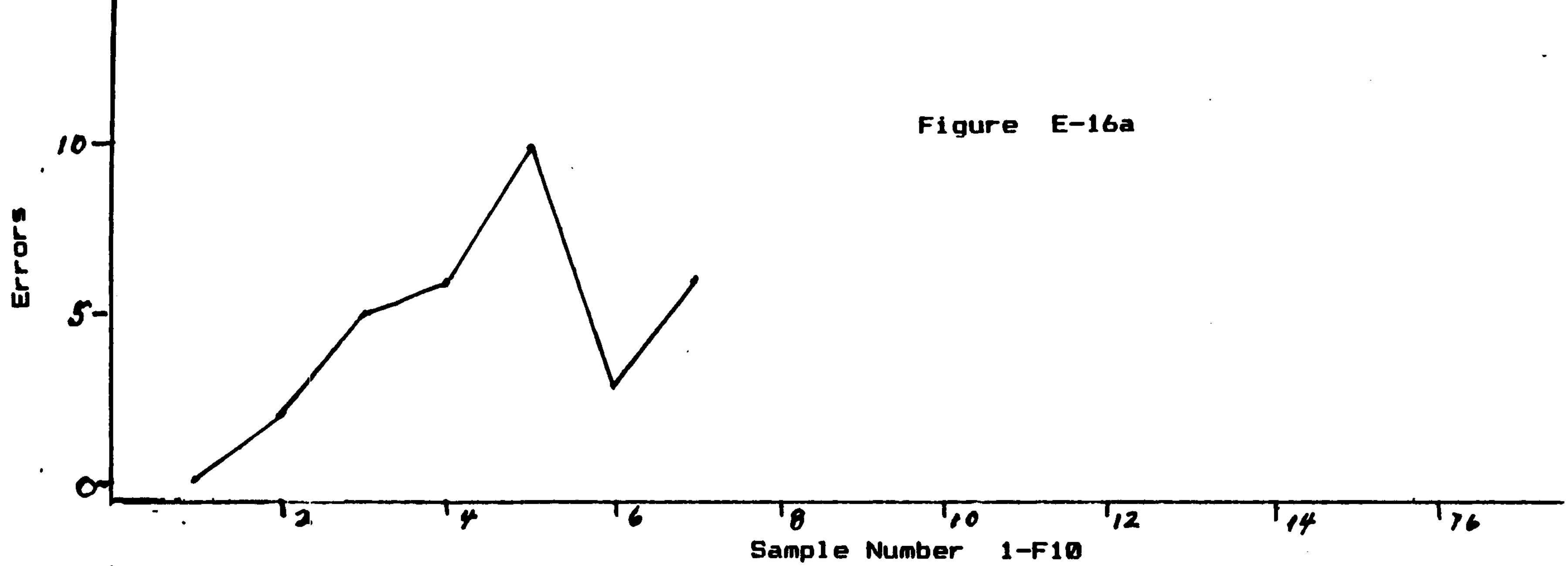


Figure E-16a

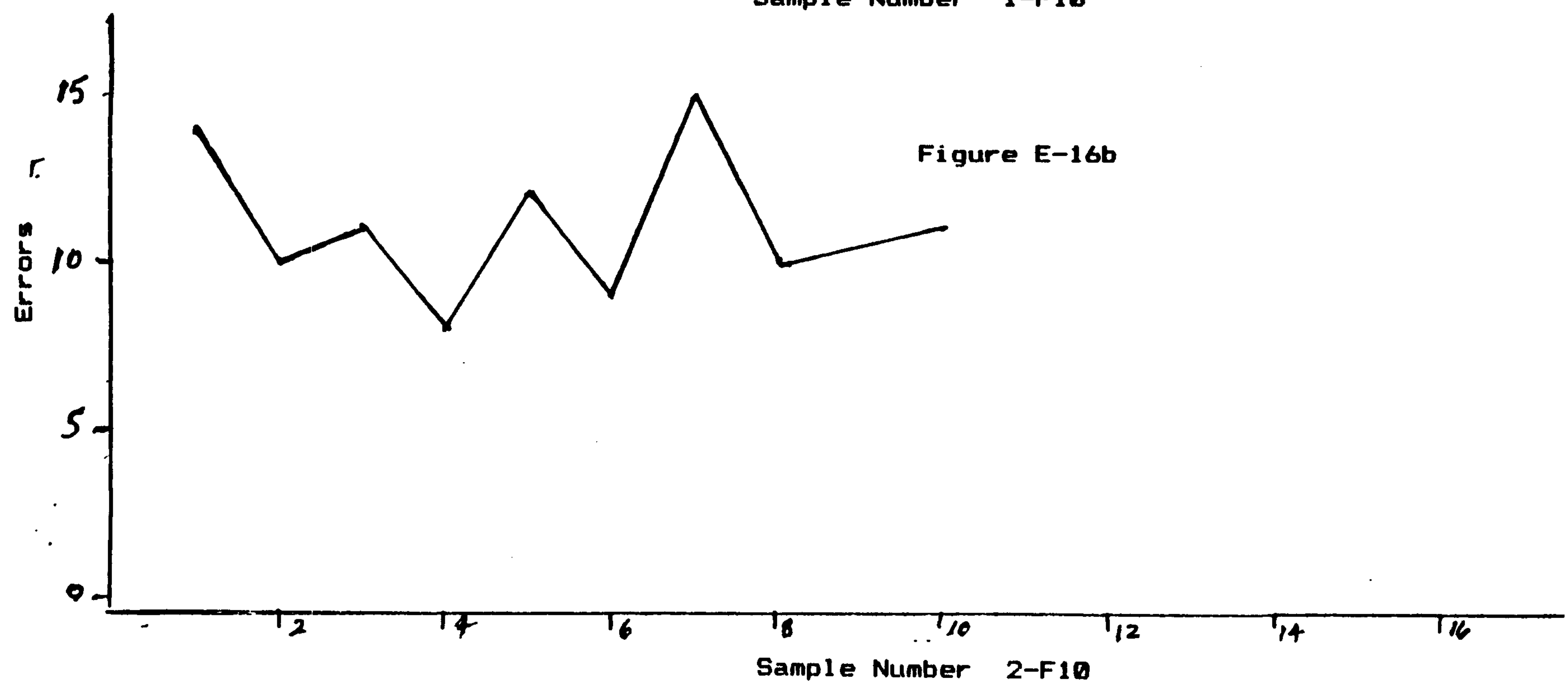


Figure E-16b

ORIGINAL PAGE IS
OF POOR QUALITY

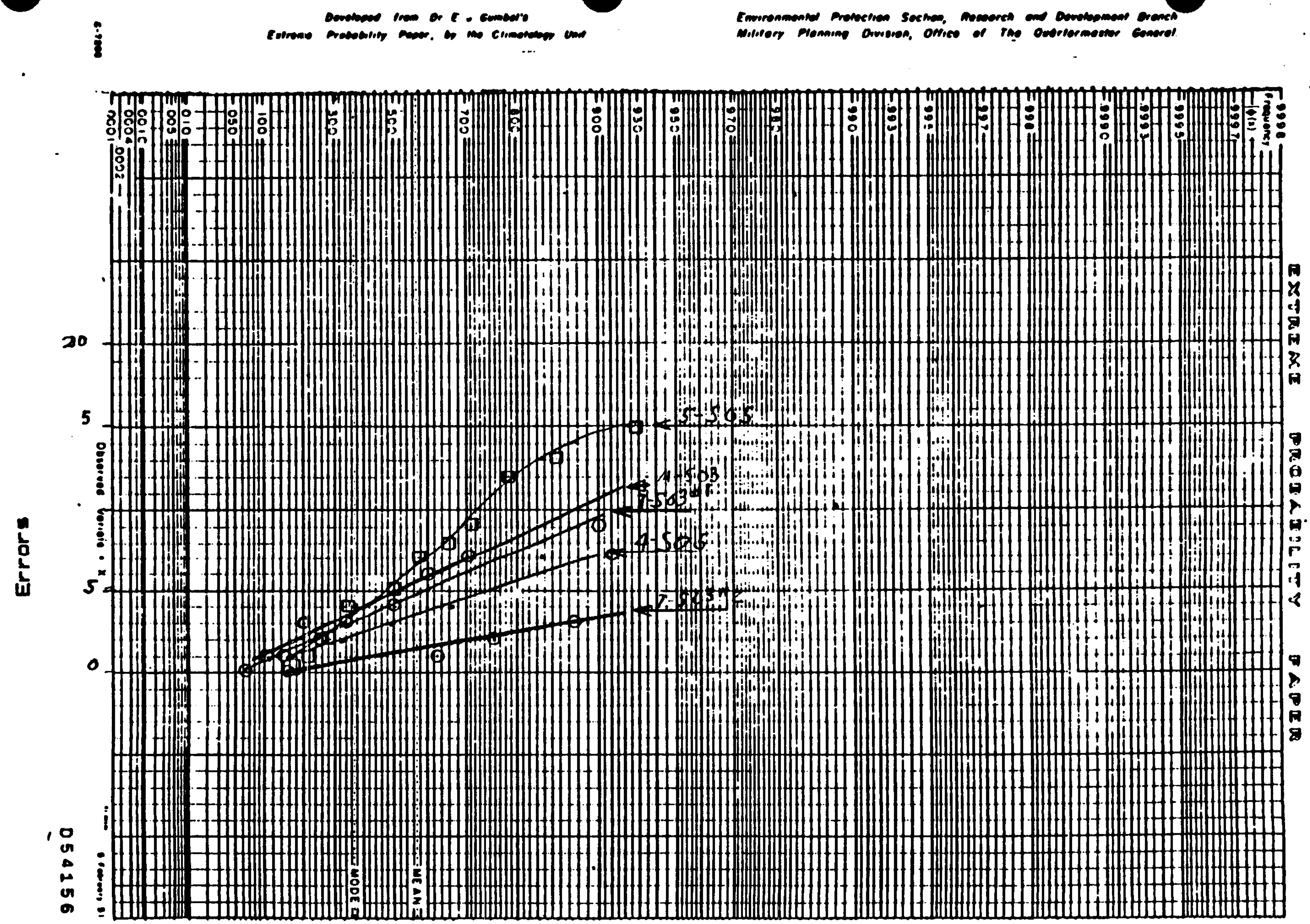


Figure E-17 Proofreading Results of Serial Number Test by VODEM

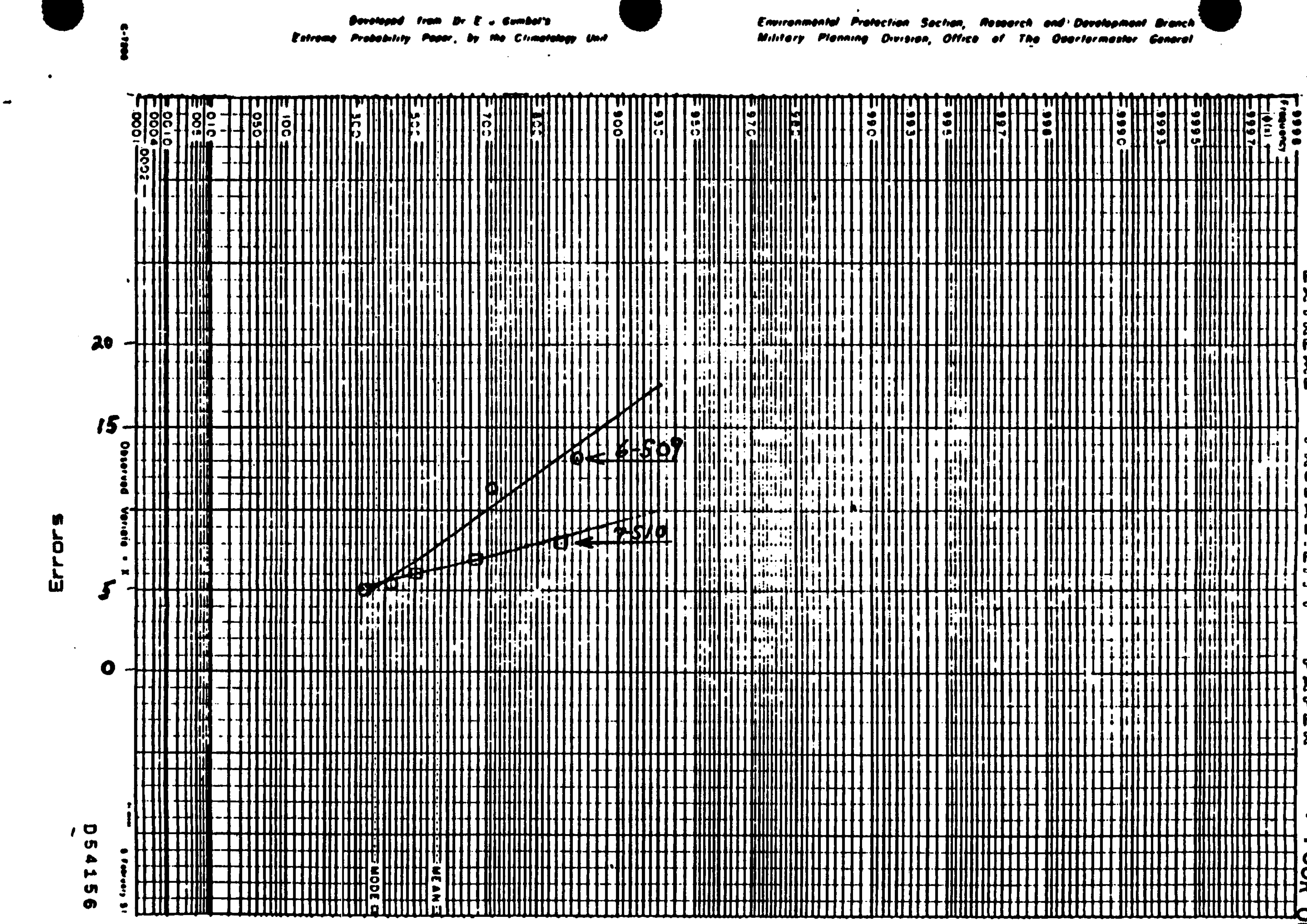


Figure E-18 Proofreading Results of Serial Number Test by STS

Individuals Precontrol Charts of Serial Number Test by VODEM

Figure E-19a

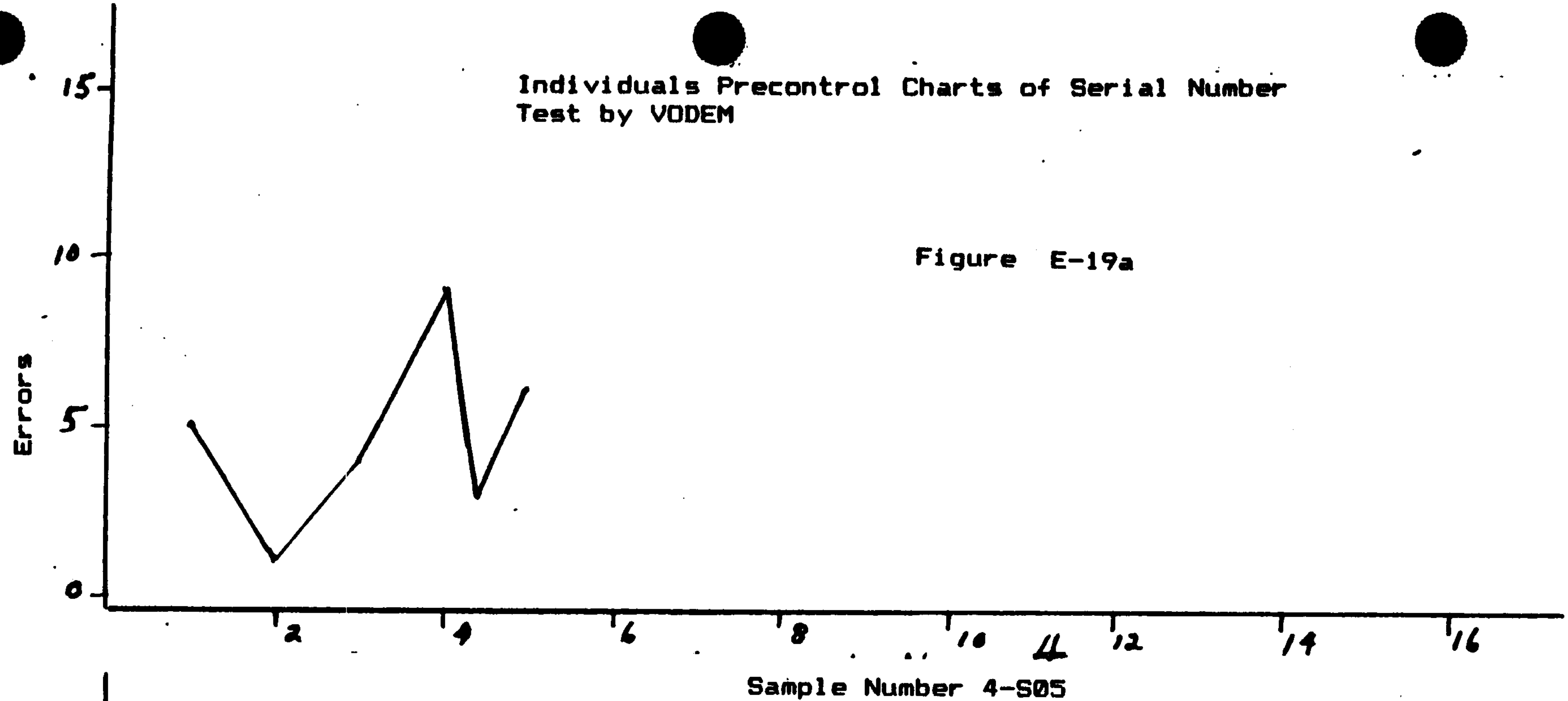
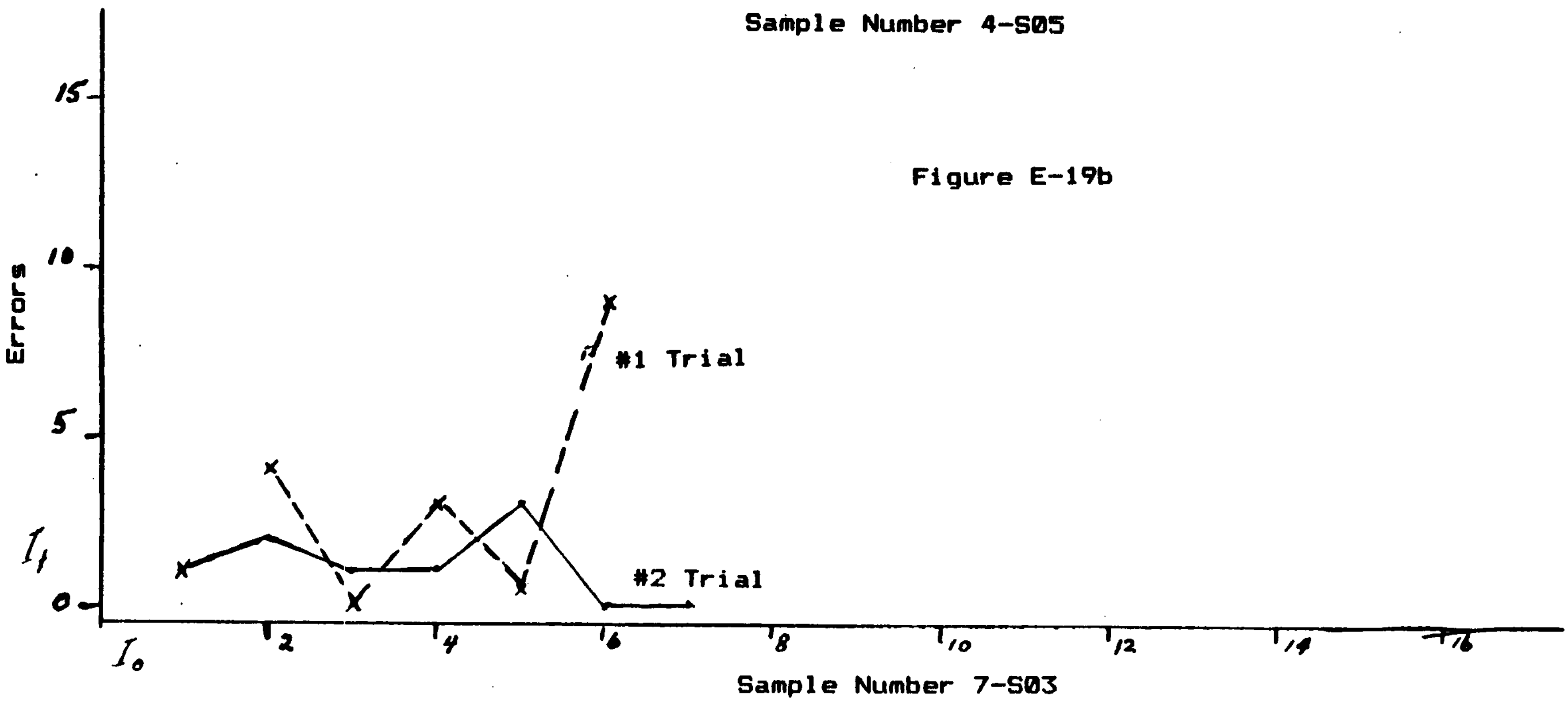


Figure E-19b



Indivi ls Precontrol Charts of Serial Number Te
by VODEM (continued)

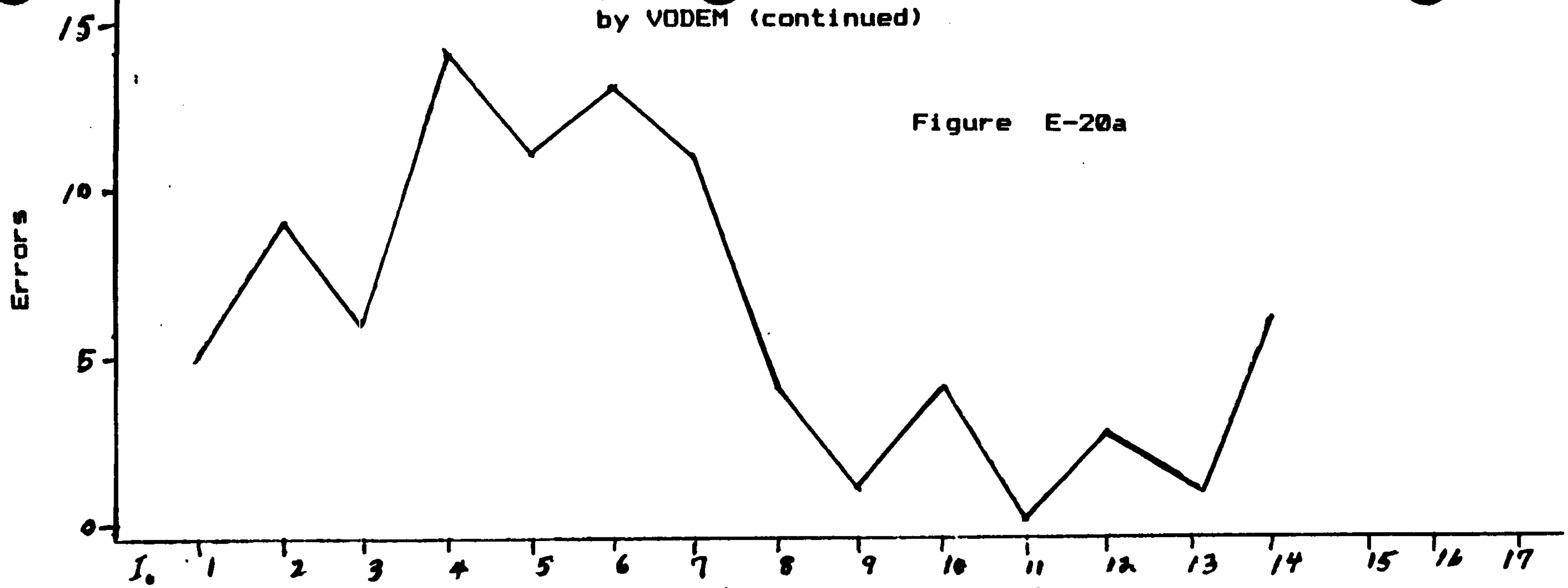


Figure E-20a

Sample Number 9-S03

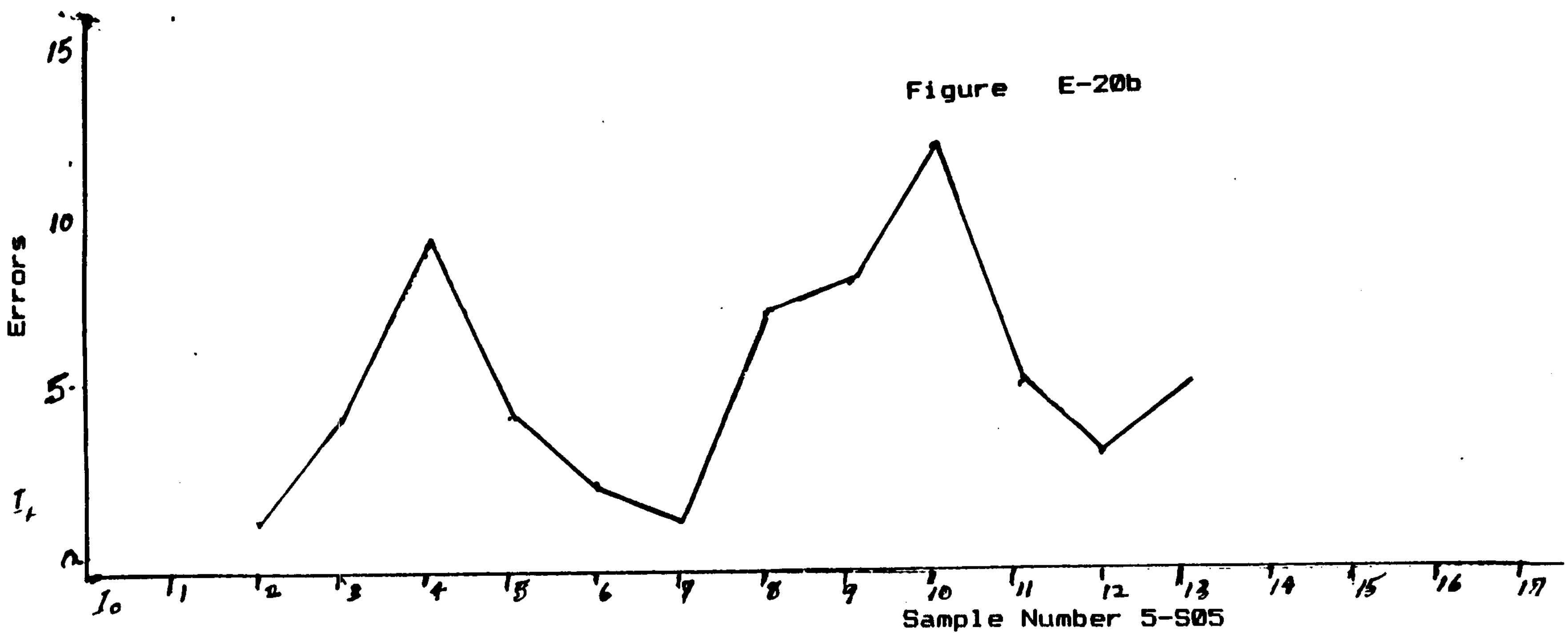
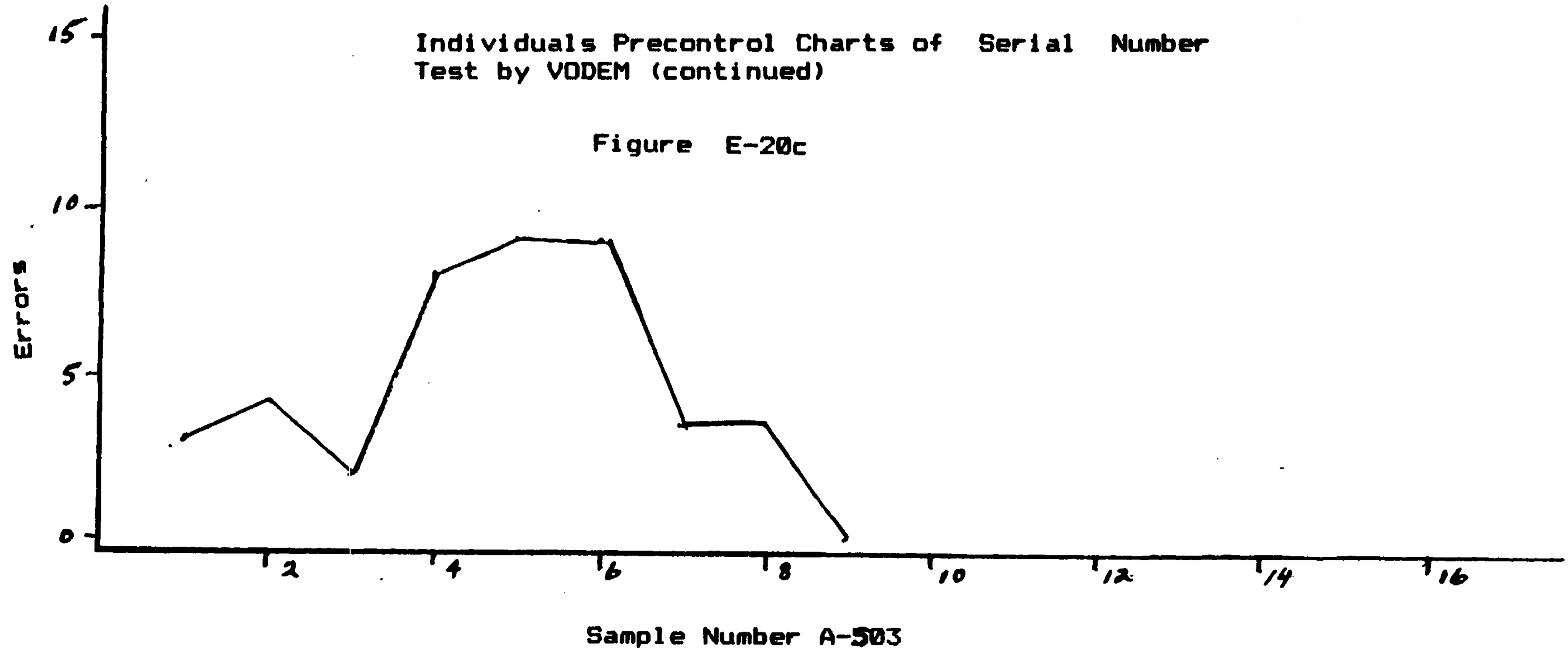


Figure E-20b

Sample Number 5-S05

Individuals Precontrol Charts of Serial Number Test by VODEM (continued)

Figure E-20c



Individuals Precontrol Charts of Serial Number Test
by STS

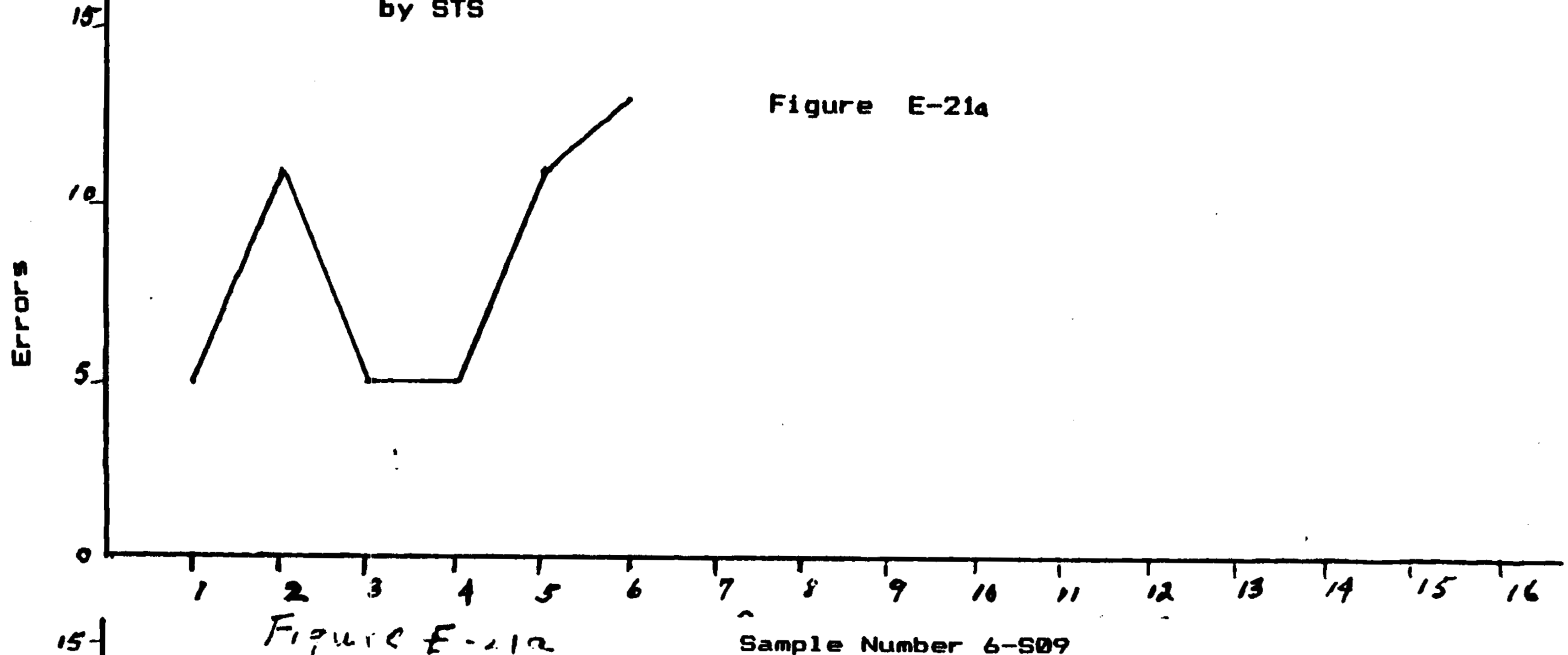


Figure E-21a

Figure E-21a

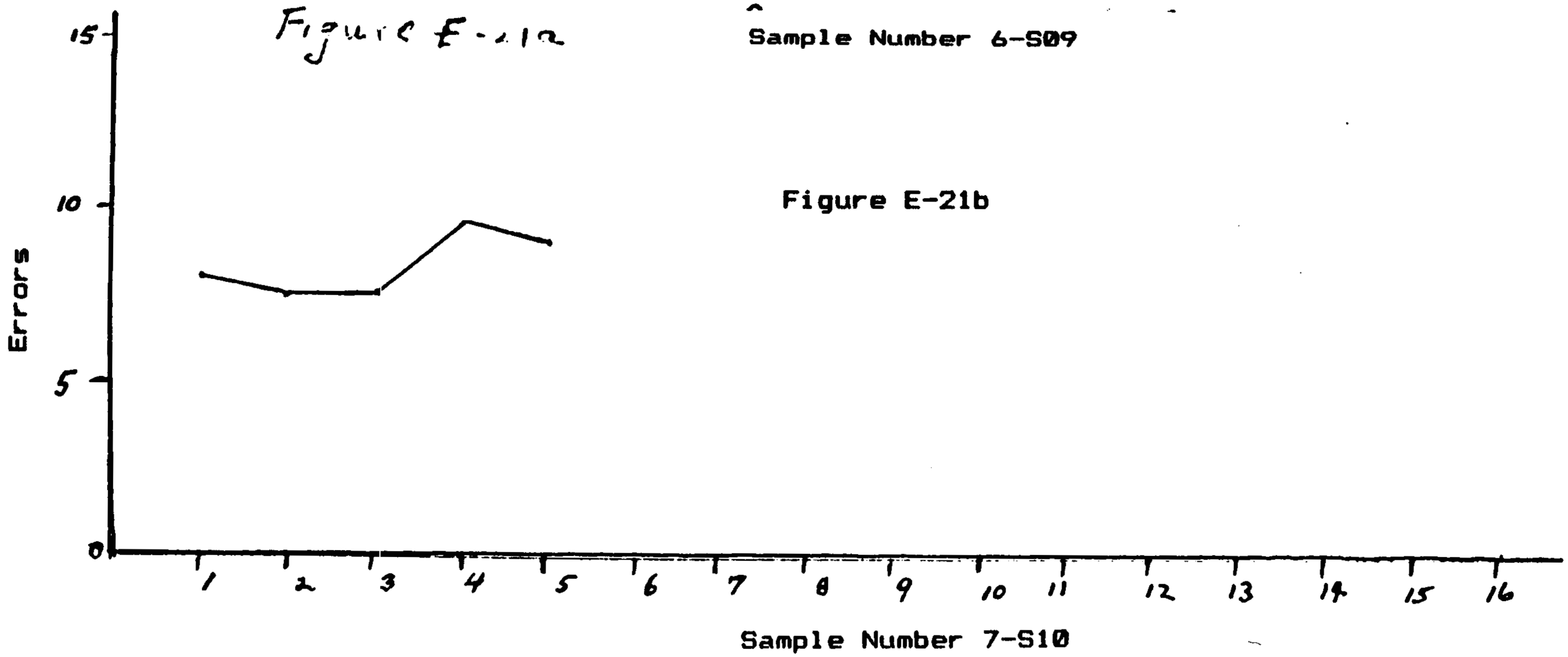


Figure E-21b

One unplanned experiment was tried. One person (7-S08) was given a serial number proofreading repeat trial by the VODEM method. In the repeat trial, this person was permitted to repeat proofreading the same material the following day. The improvement shown in both the slope of the line on the Gumbel paper (Figure E-17) and in the control chart (Figure 19b, 7-S03) is apparent. The part played in this result by experience and by memory of material is confounded. Further investigation would be required to separate the cause of the two effects.

Only two people were able to complete sufficient serial number proofreading by the STS method. Both individuals' data indicated a high error rate as indicated in Figures E-18, E-21a and E-21b.

E.5 Numerical Analysis of Extreme Probability Data

Graphic analysis by the use of Gumbel's probability paper is usually the method of choice when manual analysis must be performed. Although there is some loss in precision and accuracy the graphic method is more efficient and is satisfactory for most uses. Computer analysis, however, would permit on-line reporting. The additional costs would include development of the computer programs to allow transmission of the data to a central console or microcomputer for analysis and monitoring of work in progress.

The method would permit very early detection of differing proofreading results in time to initiate psychological investigation of the test subject for possible improvement or correction. The mathematics for incorporation into computer programs are found in Gumbel's original paper (Reference 12). Some of Gumbel's results have been reduced to simpler equations and short tables.

Table E-1 has been computed for the expected mean $\bar{y}_N$, $\bar{y}_N/\alpha$ and the expected standard deviation σ_N as functions of the number of extremes N.

Expected mean $\bar{y}_N$, and expected standard deviation σ_N as functions of the number of extremes N

Number of extremes, N	Expected mean $\bar{y}_N$	Expected mean $\bar{y}_N/\gamma$	$y_{1,N}$	Expected standard deviation σ_N	Expected standard deviation $\sigma_N\sqrt{6}/\pi$	$y_{2,N}$	$\check{y}_N$
1	2	3	4	5	6	7	8
15	0. 51284	0. 88847	2. 135	1. 02057	0. 79574	1. 476	2. 74049
20	. 52355	. 90703	2. 327	1. 06283	. 82869	1. 672	3. 02022
25	. 53086	. 91969	2. 480	1. 09145	. 85100	1. 824	3. 23855
30	. 53622	. 92898	2. 608	1. 11238	. 86732	1. 949	3. 41763
35	. 54034	. 93611	2. 718	1. 12847	. 87967	2. 056	3. 56946
40	. 54362	. 94180	2. 814	1. 14132	. 88988	2. 148	3. 70126
45	. 54630	. 94644	2. 900	1. 15185	. 89810	2. 231	3. 81767
50	. 54854	. 95032	2. 977	1. 16066	. 90496	2. 304	3. 92193
60	. 55208	. 95645	3. 112	1. 17467	. 91589	2. 432	4. 10261
70	. 55477	. 96111	3. 227	1. 18536	. 92422	2. 541	4. 25559
80	. 55688	. 96477	3. 328	1. 19382	. 93082	2. 635	4. 38823
90	. 55860	. 96775	3. 418	1. 20073	. 93621	2. 719	4. 50532
100	. 56002	. 97021	3. 498	1. 20649	. 94070	2. 795	4. 61013
150	. 56461	. 97816	3. 813	1. 22534	. 95540	3. 087	5. 01393
200	. 56715	. 98256	4. 040	1. 23598	. 96369	3. 297	5. 30082
250	. 56878	. 98539	4. 219	1. 24292	. 96910	3. 461	5. 52345
300	. 56993	. 98738	4. 366	1. 24786	. 97295	3. 596	5. 70544
400	. 57144	. 98999	4. 599	1. 25450	. 97813	3. 812	5. 99270
500	. 57240	. 99166	4. 783	1. 25880	. 98148	3. 980	6. 21561
750	. 57377	. 99403	5. 118	1. 26506	. 98636	4. 288	6. 62071
1000	. 57450	. 99529	5. 356	1. 26851	. 98905	4. 509	6. 90825

Table E-1 Expected mean y and expected standard deviation

By using the table, calculation of the expected Extreme Probability regression line is readily accomplished. Table E-2 gives an example of the way in which it would be set up for manual calculation.

Interpolation between the values given in Table E-1 is practical. The interpolations of $\bar{y}_N$ and σ_N are linear and a graphic interpolation can be made from a figure in Gumbel's paper. Equations also given are:

$$y_{1,N} = a_1 + b_1 \hat{y}_N$$

$$y_{2,N} = a_2 + b_2 \hat{y}_N$$

Equations for the calculation of the probabilities of are also given:

$$\Phi y_{1N} = \bar{y}_N/\alpha \quad ; \quad \Phi y_{2N} = \sigma\sqrt{\frac{6}{\pi}}$$

$$\hat{y}_N = N/N+1$$

Additional calculations are defined in Table E-3.

Data Series	N	$\bar{X}$	S_x Normal	Y_N *	σ_N *	$\frac{1}{\alpha}=\frac{S_x}{\sigma_N}$	$u=\bar{x}-\frac{\bar{y}_N}{\alpha}$	$X=\mu+\frac{y}{\alpha}$
4-F08	15	1.5333	1.1125	.5184	1.0257	1.09008	.9684	.9482 + 1.09008 y

where -

- * is taken from Table E-1
- n number of error in a sample
- $\bar{X}$ mean
- S_x standard deviation (normal) of the sample
- $1/\alpha$ slope of the line on Gumbel's probability paper
- u the mode
- μ the mean (theoretical)

Table E-2 Example: Calculation of Extreme Probability Chart Regression

Characteristic	Largest value	Smallest value	Normal variate

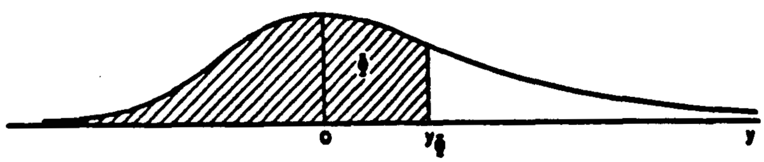
A. *Distribution*

1. Cumulative probability function for reduced variate, Φ	$e^{-e^{-y}}$	$1-e^{-e^{y}}$	$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$
2. Reduced variate	$y=\alpha(x-u)$	$y=\alpha(x-u)$	$z=\frac{1}{\sigma}(x-\mu)$
3. Original variate x, in terms of reduced variate	$x=u+y/\alpha$	$x=u+y/\alpha$	$x=\mu+\sigma z$

B. *Reduced variate y*

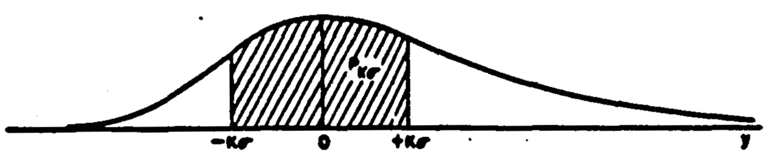
4. Moment generating function $G(t)$	$\Gamma(1-t)$	$\Gamma(1+t)$	$e^{\frac{1}{2}t^2}$
5. Mode $\tilde{y}$	0	0	$\tilde{z}=0$
6. Mean $\bar{y}$	$\gamma=0.57722\cdots$	-0.57722	$\bar{z}=0$
7. Median $\breve{y}$	$-\lg\lg 2=0.36651$	-0.36651	$\breve{z}=0$
8. Standard deviation σ	$\pi/\sqrt{6}=1.28255$	1.28255	1
9. Skewness β_1	1.29857	-1.29857	0
10. Kurtosis β_2	27/5	27/5	3

C. *Deviates y_Φ corresponding to given probability points Φ*



11. $\Phi=.95$	2.97020	1.09718	1.64485
12. $\Phi=.99$	4.60015	1.52718	2.32635

D. *Area $P_{K\sigma}$ included within mode $\pm K\sigma$*



13. $K=1$; $P_{K\sigma}=\Phi(\sigma)-\Phi(-\sigma)$	0.73064	0.73064	0.68269
14. $K=2$; $P_{K\sigma}=\Phi(2\sigma)-\Phi(-2\sigma)$	0.92597	0.92597	0.95450
15. $K=3$; $P_{K\sigma}=\Phi(3\sigma)-\Phi(-3\sigma)$	0.97890	0.97890	0.99730
16. Variate y for which $P_y=0.68269$	1.14071	1.14071	1.00000
17. Variate for which $P_y=0.95450$	3.06685	3.06685	2.00000

E. *Original variate, x*

18. Mode $\tilde{x}$	u	$-u$	μ
19. Mean $\bar{x}$	$u+\gamma/\alpha$	$-u-\gamma/\alpha$	$\mu+\sigma\bar{z}=\mu$
20. Median $\breve{x}$	$u+0.36651/\alpha$	$-u-0.36651/\alpha$	$\mu+\sigma\breve{z}=\mu$
21. Standard deviation σ_x	$\frac{1}{\alpha}\cdot\sigma=\frac{1}{\alpha}\cdot\frac{\pi}{\sqrt{6}}$	$\frac{1}{\alpha}\cdot\sigma=\frac{1}{\alpha}\cdot\frac{\pi}{\sqrt{6}}$	$\sigma\cdot 1$

Table E-3 Extreme Value Calculations and Comparison with the Normal Variate

From the table, the meaning of the first three boxes A, B, and C are clear. The first three lines of box D show the probabilities of $\pm\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ about the mode. The last two lines (Section D) give the reduced variates of y, for which the probability (P) is represented by the area under the extreme value frequency distribution function Φ (y) bounded by the symmetric interval which has the same value as that for the normal curve corresponding to $\pm\sigma$ and $\pm 2\sigma$. The last box E shows the relationship between the important characteristics of the original variate x and the reduced variate y for the asymptotic distribution.

The expected mean and standard deviation of reduced extreme values have already been calculated for N between 20 and 100 and are presented in Table E-4. This table is extremely valuable for manual or computer calculation of control charts.

Expected means and standard deviations of reduced extremes

N	$\bar{y}_N$	σ_N	N	$\bar{y}_N$	σ_N
20	0.5236	1.0628	50	0.5485	1.1607
21	.5252	1.0695	51	.5489	1.1623
22	.5268	1.0755	52	.5493	1.1638
23	.5282	1.0812	53	.5497	1.1653
24	.5296	1.0865	54	.5501	1.1667
25	.5309	1.0915	55	.5504	1.1681
26	.5320	1.0961	56	.5508	1.1696
27	.5332	1.1004	57	.5511	1.1708
28	.5343	1.1047	58	.5515	1.1721
29	.5353	1.1086	59	.5518	1.1734
30	.5362	1.1124	60	.5521	1.1747
31	.5371	1.1159	62	.5527	1.1770
32	.5380	1.1193	64	.5533	1.1793
33	.5388	1.1226	66	.5538	1.1814
34	.5396	1.1255	68	.5543	1.1834
35	.5403	1.1285	70	.5548	1.1854
36	.5410	1.1313	72	.5552	1.1873
37	.5418	1.1339	74	.5557	1.1890
38	.5424	1.1363	76	.5561	1.1906
39	.5430	1.1388	78	.5565	1.1923
40	.5436	1.1413	80	.5569	1.1938
41	.5442	1.1436	82	.5572	1.1953
42	.5448	1.1458	84	.5576	1.1967
43	.5453	1.1480	86	.5580	1.1980
44	.5458	1.1499	88	.5583	1.1994
45	.5463	1.1519	90	.5586	1.2007
46	.5468	1.1538	92	.5589	1.2020
47	.5473	1.1557	94	.5592	1.2032
48	.5477	1.1574	96	.5595	1.2044
49	.5481	1.1590	98	.5598	1.2055
			100	.5600	1.2065

Table E-4 Expected Means and Standard Deviation of Reduced Extreme

E.6 Conclusions

The most valuable contribution of statistical evaluation in this program is the discovery of the fact that, by at least two modes of proofreading and two groups of subjects, errors in proofreading follow Gumbel's extreme value theory. The error distribution of all methods may follow the same distribution. This discovery makes the use of statistical control charts possible both manually and by computer on-line. Secondly, it also demonstrates the need for some training in order to perform proofreading satisfactorily by any method. Finally with computer-aided proofreading methods, particularly with on-line computerized analysis, an objective means of measuring and obtaining the required accuracy for communication between networked computers and between data entry terminals and centralized data bases is now possible. Another possible application would be the objective selection of personnel for accurate and economical proofreading.

Section F

HUMAN ERROR MODEL

F.1 The Error Decay Theory

Let us consider the idealization of the experiments reported in the previous sections, somewhat like a "thought" experiment. We describe the experiment in terms of an observer far removed from the site so that he is unable to observe any of the psychological, biological, and motor mechanisms that the subject uses to perform the detection, indication, and correction processes of the proofreading experiments. The information available to the observer include the PDP(1) and PDP(2) lines shown in Figure C-1. He is able to determine the state of each PDP pair:

- State I - an error in PDP(2)
- State II- no error in the PDP pair

Occasionally a transition will occur in a PDP pair of characters: an error in a PDP pair in State I is (somehow) corrected and becomes error-less and the PDP pair of characters are in State II. Thus, in the transition process errors become non-errors. By measuring some "time" parameter, the observer could describe this process as an exponential decay process, somewhat similar to the theory of radioactive decay of atoms. The decay of errors into non-errors is an <u>error decay process</u>.

We turn now to the formulation of an Error Decay Theory using the results of the "thought" experiment. Consider a single character error. Decay can only occur when the operator's visual field passes over it. This time interval is the proofreading exposure time window, Δt. The probability of error decay is assumed proportional to this imaging time:

$$P_{decay} = \lambda \Delta t \quad \text{, where } \lambda \text{ is the decay constant.}$$

The probability that the error does not decay is:

$$1 - P_{decay} = 1 - \lambda \Delta t$$

If the error survives this time interval in the first proofreading cycle, it will not have a chance to decay until the second proofreading cycle, q=2, when the proofreading window passes over it again. The probability that it does not decay in the second proofreading cycle, q=2, is also $(1 - \lambda \Delta t)$. The probability that the error survives q=1 and q=2 is:

$$(1 - \lambda \Delta t)(1 - \lambda \Delta t) = (1 - \lambda \Delta t)^2$$

For q such proofreading cycles, the probability of survival is

$$(1 - \lambda \Delta t)^q$$

If the total imaging time per error is $t^* = q \Delta t$, the probability of survival is:

$$(1 - \lambda t^*/q)^q$$

where the asterisk is to remind the reader that t^* is not a continuous time record, but the sum of the Δt's over the proofreading cycles.

Having made the point that t is not a continuous time record, we drop the asterisk from the notation. The probability that the one error survives after the imaging time of t is the limit of the above expression as the number of cycles q becomes very large:

$$\lim_{q \to \infty} \left(1 - \frac{\lambda t}{q}\right)^q = e^{-\lambda t}$$

The above result for single error decay can be generalized to a large number of errors. If initially there are a large number of errors, N_o, the fraction of errors remaining at imaging time t is:

$$\frac{N(t)}{N_o} = e^{-\lambda t}$$

F.1.2 Measurable Parameters

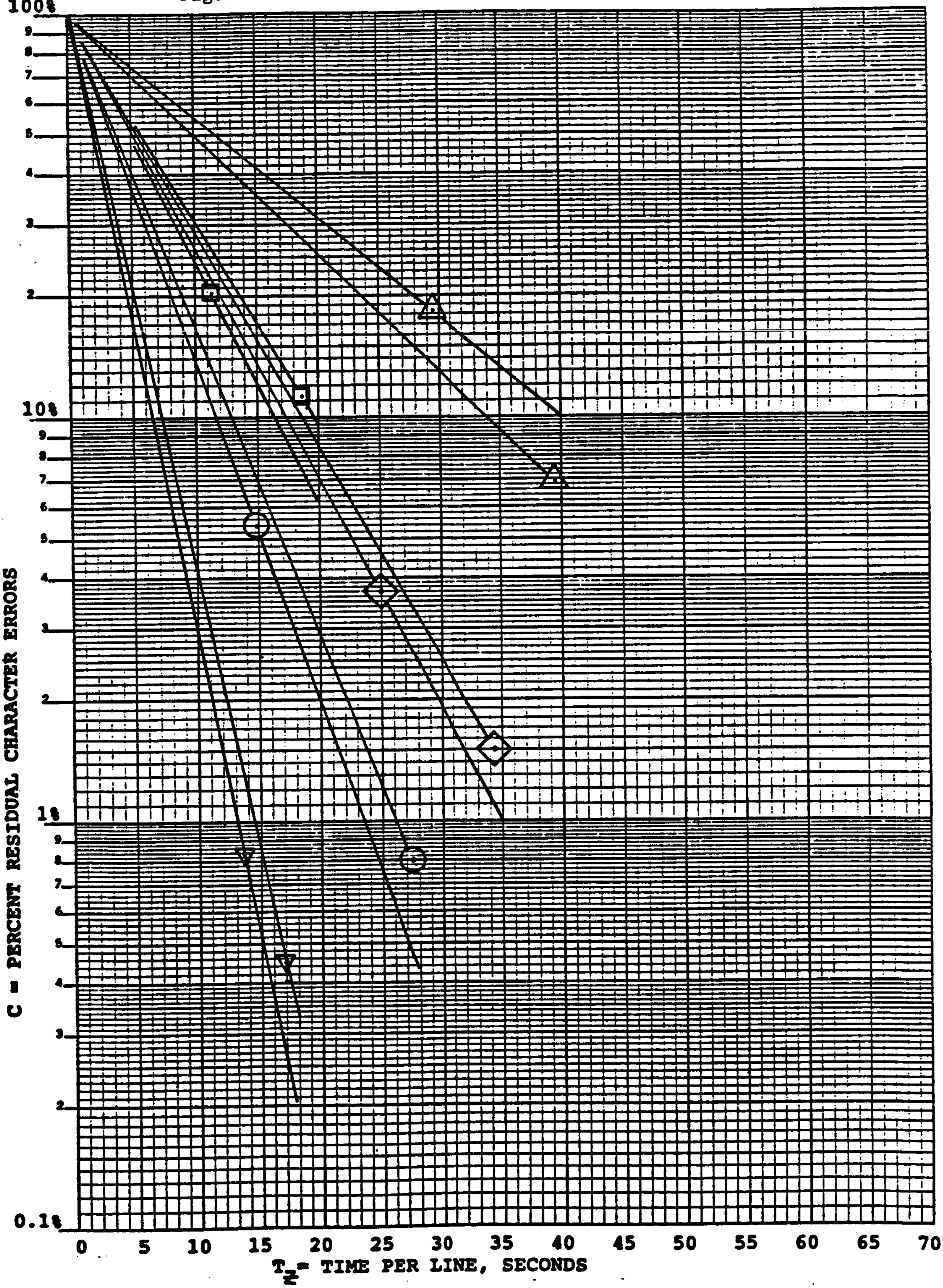
In the development above the imaging time t is not measurable for the experiments of the Phase I study. However, an approximation to the imaging time is T_z, the proofreading time for lines without error, which is measured in the experiments. The fraction of errors remaining is also measured as C, which is the percent residual character error. This allows the determination of the decay constant.

F.1.3 Experimental Evidence for the Error Decay Model

The complete development of the Error Decay Model is a task for the Phase II study. However, one subject from the previous study of handicapped individuals and all four subjects from Study B of the current contract showed some evidence that the proofreading experiments exhibit error decay characteristics. For these individuals the same proofreading file (same error density) was continued over two test sessions. For the handicapped individual, a word file at VODEM conditions at the middle error density was used. For the subjects from Study B, the file was Serial T03 at the VODEM condition. Figure F-1 shows that the second test session for each individual gave a faster T_z proofreading time and a higher error rate. (The upside-down triangle was for the handicapped individual and the other symbols follow the notation of Table D-1.)

The Error Decay Model, when sufficiently validated in Phase II, would provide a very valuable means of describing human error for proofreading tasks. It would be expected that the decay constant would assume different values, depending on the difficulty of the proofreading material. There is also the possibility that the imaging time can be indirectly controlled through control of display parameters at a work station.

Figure F-1. Evidence for Error Decay Behavior



F.2 Auxiliary Observations

The error extraction data reported here represents tens of thousands of perceptual and motor events. The basic behavioral data, recordings of key strokes in controlled stimulus and procedural contexts, number in the thousands. The analysis reported above of the error correcting incidents gives a very fine-grained depiction of the detection and overlooking of textual mismatches, fairly precisely located in time. The theoretical framework, which utilizes both rate statistics and direct estimates of the λ parameter, turns out to have provided a sensitive method of summarizing changes in the cognitive error detection processes.

These quantitative, behavioral observations give us information that the participants themselves may be unaware of, on their response to differing error densities and conditions of presentation, etc. All the same, the precise behavioral techniques do not tell us all that is going on in the situation. People may be applying strategies of error detection, feeling one way or another about the task, and having thoughts about matters both relevant to the assigned tasks and to other aspects of the situation that they are in.

To gain clues about the feelings and cognitive strategies of the participants, auxiliary information was sought. Participants rated the proofreading sessions according to *ease or difficulty* and *how tiring* they were, on 5 inch continuous horizontal line scales, anchored at the endpoints and at the center, "medium" or "average" point. In addition, post-participation discussions were held individually with several persons from Study A, and with the participants in Study B as a group.

Ratings of tasks. On the average, the proofreading tasks were rated as fairly easy. Means for detecting errors in Random Words, under various conditions, were a little less than 1 on a scale of 5, where 0 would be maximally easy. (Clearly, the tasks were not so easy as some participants thought, or at least indicated on the rating scales.)

For Vodem presentation of Serial Numbers, the low error density condition was rated in the sample of Study B as very slightly easier than the middle density condition (1.36 vs. 1.52, not statistically significant). The same trend was observed in Study A (1.32 vs. 1.85, but with different sets of subjects). Study B found the STS condition less easy in the low density condition (mean = 2.22), as would be expected, but about as easy as the Vodem presentation of middle density error material.

Tasks overall were rated as moderately tiring, in the 1 to 2.5 area on a scale of 0 to 5, where 0 would be maximally tiring. One might expect that STS would in general be rated as more tiring than Vodem presentation. Sometimes it was and sometimes it wasn't. The most comparable data, from Study B, indicated that at medium error density, subjects found Vodem presentation a little more tiring than STS (1.69 vs. 2.29). At low error densities, STS was rated as more tiring (2.42 vs. 1.58).

Because of small sample size, and the multiple meanings that rating scale labels sometimes have for study participants, these statistics should be interpreted cautiously.

Some individual patterns in rating the tasks may be noted. Perhaps coincidentally, the single best performers in Study A and Study B had similar and distinctive rating patterns. They rated the tasks as moderately easy but as much less tiring than did most other participants. Two or three participants in Study A made clear by their comments that they found the tasks aversive (especially Serial Numbers). They characteristically rated the sessions as maximally tiring but also as maximally easy. Interestingly, performance of these subjects varied from what seemed to be rushed carelessness to thorough and slow extraction of almost every error.

Descriptions of proofreading process. Discussion by participants of their view of the process revealed much qualitative information, which will not be reported in detail here. Unfortunately, it was not possible to discuss the subjective side of the process with the participants from Study A who had most experience in side-to-side proofreading as parts of their jobs, owing to time constraints. We are especially curious to do so, since some of the more experienced persons had performance patterns that departed from straightforward predictions. (An analysis of individual protocols, comparing each predicted relationship between pairs of sessions, showed that *all* the reversals of the prediction that STS would be less effective than Vodem presentation came from participants with extensive work experience in proofreading, much of which was probably in side-to-side mode.)

Participants in Study B volunteered several individual strategies, including "scanning between the lines" rather than doing up-down comparisons in Vodem mode; or using ABBA patterns of comparison in STS mode rather than back-and-forth ABAB patterns. Many of their comments reflect processes that are understood well, such as the "chunking" or grouping of letters and numbers into small sequences.

Stretches of text with few discernible errors were regarded as difficult to deal with (no matter what ratings were given them). The lines "ran together".

Participants differed in strategies for dealing with high error densities. Some said they were careful to rescan lines after catching one error; others moved on. Several mentioned what are discussed as "betting" phenomena in the psychological literature: either anticipating errors when none had been detected recently; or conversely betting that if an error were found, another would be close by.

It was difficult to relate the various comments to actual performance level (really a comparative matter between individuals), but the discussion taken as a whole did suggest that the error detection process was not entirely a passive program for these individuals. The qualitative information obtained from these discussions should be useful in the future, in coming to identify individual styles, so that effective feedback can be given to help adjust error rates.

Section G

CONCLUSIONS

This Phase I study has demonstrated through experiments with clerical volunteers and paid subjects that the VODEM Technology can provide significant gains in data accuracy for manual data entry tasks.

A Human Error Model has been developed, using the exponential decay theory to describe error rates during proofreading. With further development in Phase II the two-parameter theory of decay constant and imaging time should permit the determination of error rates.

The diagnostic computerized procedures to evaluate personnel have been developed as input to both the experimental program and the statistical evaluation.

The most valuable contribution of statistical evaluation in this study was the demonstration that errors in proofreading follow Gumbel's extreme value theory. This allows the development of statistical control charts to control task accuracy.

With the successful completion of Phase I, it has been demonstrated that techniques have been developed in error decay theory and statistical analysis to allow an objective means of measuring and obtaining the required data accuracy for communication between networked computers and between data entry terminals and centralized data bases.

Section H

REFERENCES

1. Woo, John, Jr., and Lien, Edmond, "Electronic Rule for Precise Distance Measurement and Distance Setting," U. S. Patent 4,158,229, June 12, 1979 (Canadian Patent 1091439 and United Kingdom Patent GB 2036394B).
2. Rhodes, Wayne L., "The Disproportionate Cost of Data Entry," Infosystems, October 1980, pp.70.
3. Bailey, Robert W., Human Error in Computer Systems, Prentice Hall, NJ, 1983, pp.29.
4. Cardenas, A. F., "Data Entry: A Cost Giant," J. of Systems Management, Aug. 1973, pp.39.
5. Woo, John, Jr., Nocke, H., Hays, D., Wilson, H., and Scott, N., "Development of a Computer-Assisted Proofreading System for the Handicapped," Gamma Research, Inc. Final Technical Report (Phase I SBIR), prepared for U. S. Department of Education, Contract 400-83-0035, Office of Special Education and Rehabilitative Services, Contract Monitors Mr. Joseph Traub (202-732-1189) and Dr. Edward Esty (202-254-8247), July 27, 1984.
6. Van Cott, H. P., and Kinkade, R. G., Human Engineering Guide to Equipment Design (Revised Edition), McGraw-Hill, New York, 1972, pp. 355.
7. Klemmer, E. T., and Lockhead, G. R., "Further Data on Card Punch Operator Performance," Research Note No. NC 39, Thomas Watson Research Center, Yorktown Heights, NY, 1962.
8. Minton, George, "Inspection and Correction Error in Data Processing," Journal of the American Statistical Association, Vol. 64, 1969, pp. 1256-1275.
9. Naus, Joseph, Data Quality Control and Editing, Marcel Dekker, NY, 1975.
10. GSA, General Services Administration, General Procurement Branch, 75 Spring Street Atlanta, GA, Solicitation Number AT/TC 19912, 1986.
11. Elderton, W. P., Johnson, N. L., Systems of Frequency Curves, Cambridge University Press, 1969.
12. Gumbel, E. J., "Statistical Theory of Extreme Values and Some Practical Applications," NBS Applied Mathematics Series 33, U. S. Government Printing Office, Washington, DC, February 12, 1954.
13. Western Electric Co., Inc., Statistical Process Quality Control Handbook, 4th printing, September 1970.